Dr. Homi Bhabha and Cosmic Ray Research in India

(B.V. Sreekantan, National Institute of Advanced Studies, Bangalore)

Dr. Homi Bhabha is well known among the scientists of India and the public at large as the Father of India's Atomic Energy programme. What is not so well known however, particularly among the younger generation is the fact that he made outstanding contributions in the field of theoretical physics and played a very important role in initiating and fostering cosmic ray research in India. Thanks to the unique start given by him in the 40's, cosmic ray research in India grew into one of the largest activities in the world covering all aspects of the radiation in the 50's and 60's. In recognition of this, the Cosmic Ray Commission of the IUPAP held its 7th International Conference on Cosmic Rays, one of the earliest in the series at Jaipur in India in 1963. It is most interesting and instructive to know how all this was achieved and what exactly motivated a young man in his thirties to initiate work in a highly sophisticated and highly technology dependent field so early in India. Let us look at the background of Dr. Bhabha and the times in which he began his research.

Homi Bhabha at Cambridge

Homi Bhabha was born on 30th October 1909 and had his school education at the Cathedral and John Connon High School in Bombay and his college education at the Elphinstone College and the Royal Institute of Science. At the age of 18, he left for England to pursue further studies at

Cambridge. As desired by his parents, he completed his Mechanical Tripos with distinction and pursuaded them to let him do a Mathematical Tripos since his own interests were in physics. Immediately after his second Tripos, he got a travelling fellowship and had the wonderful opportunity of working for short periods with Wolfgang Pauli at Zurich and with Enrico Fermi at Rome. In 1934, he was awarded the Issac Newton Studentship at Cambridge which enabled him to complete his Ph.D under R.H Fowler. He continued his research at Cambridge till 1939. The research that he did during this period had a direct bearing on the resolution of several important issues of cosmic ray phenomena and the interactions of particles especially electrons, protons and photons at high energies, in the context of the developments in the field of quantum mechanics and relativity. To appreciate these contributions of Bhahbha it is necessary to become familiar with the status of cosmic ray studies in the early 30's.

Cosmic rays in the atmosphere: The `Soft' and `Penetrating' Components

The presence of a penetrating ionizing radiation of extraterrestrial origin was established in 1912 by Victor Hess through a series of manned balloon experiments. The name `Cosmic Rays' was given to this radiation by Millikan in 1925.

Analysis of the radiation at Sea Level and mountain altitude by a series of experiments with Geiger-Muller telescopes and magnetic cloud chambers, revealed that the radiation comprised two components with distinctly different properties. One component called the `soft component' was easily absorbed in a few centimeters of lead, quite frequently multipled in number in passing through thin sheets of lead and also arrived at the observational level in multiples - as a shower of particles separated by several tens of centimeters. The second component, called penetrating component could penetrate large thicknesses of matter even a meter of lead, without multiplying. The only fundamental particles that were known at that time were the electrons, the photons, the protons and the alpha-particles. To this small list two more were added in 1932, the Neutron and the Positron. The position was discovered by Anderson in a cloud chamber that had been set up to analyse the cosmic ray beam. The positron discovery was a great triumph to the relativistic quantum mechanical formulation of the theory of the electron achieved by Dirac at Cambridge. Around the same time in the early 30's, Blackett and Occtialini who were also working at Cambridge had recorded several instances of multiple charged particles which had obvious non-ionizing links between pairs of them. These events fitted beautifully the phenomenon of pair production of quanta into electron-positron pairs according to Dirac's theory. The calculations on the energy losses of charged particles by Bethe and Heitler revealed several surprising features - higher losses for lighter particles, (more for electrons than for protons of the same energy), higher losses in passing through matter of higher atomic number, and higher losses at higher energies.

All these features came in very handy in the explanation of cosmic ray anomalies. Clearly, Bhabha was at the right place at the right time. The very first paper of Bhabha entitled "Zur Absorption der Hohenstrahlung" published in 1933 in Zeitschrift Fur physik was concerned with the explanation of the absorption features and shower production in cosmic rays. In 1936, Bhabha in collaboration with Heitler formulated the cascade "theory of the electron" according to which a high energy electron passing through matter gave rise to a high energy photon by bremstrahlung process and the photon in turn produced a pair of positive and negative electrons; these in turn led to further production of photons and the cascade process continued until the energy of the particles fell below a critical value. Carlson and Oppenheimer also developed a similar theory simultaneously in the USA. Based on Bethe-Heitler cross sections. Bhabha and Hietler made quantitative estimates of the number of electrons in the cascade at different depths, for different initiating energy of the electrons. These calculations agreed with the experimental findings of Bruno Rossi in cosmic ray showers. The problem of the `soft' component was thus totally resolved.

In a classic paper entitled "on the penetrating component of cosmic radiation " communicated to the proceedings of the Royal Society in July 1937, Bhabha made a careful analysis of the experimental data on the soft and penetrating components and concluded that a `breakdown' of the quantum mechanical theory of radiation at higher energies as proposed by some theorists would not explain the experimental results on the latitude effect of cosmic rays and the shape of the transition curve of large cosmic ray bursts. He emphasized that these features would find a natural explanation if cosmic radiation contained charged particles of mass intermediatic between electron and proton masses and set the mass as ~100 electron masses.

Around the same time, Neddermeyer and Anderson and Street and Stevenson discovered in their cloud chamber experiments, charged particles of intermediate mass whose mass was set at ~ 200 electron masses. The name `meson' was given to this new particle.

In a paper in Nature in 1938, Bhabha predicted that this meson would be unstable and would probably decay into an electron and neutrino. The phenomenon of meson decay was helpful in resolving anomalous absorption of the penetrating component in the atmosphere. The relativistic elongation of time as predicted by the special theory of relativity was confirmed through meson decay experiments.

Dr. Bhabha and Cosmic Ray Research at the Indian Institute of Science, Bangalore

Bhabha came on a brief holiday to India in 1939. He could not go back to England as planned, since the second world war broke out in September 1939, and there was the prospect of heavy bombing over England by the Germans. Bhabha decided to stay back in India for a while. This decision turned out to be a turning point, a landmark not only in the academic career of Bhabha, but also in the advancement of Indian Science and Technology in the post independent era.

Bhabha joined the Physics Department of the Indian Institute of Science, headed by C.V. Raman. He got a special grant from the Sir Dorabji Tata Trust. He gathered some students to work with him in theoretical particle physics and one of them was Harish Chandra, who later held a professional chair in mathematics at the Princeton Institute of Advanced Studies.

In parallel, Bhabha also started experimental work in cosmic rays. He was cognisant of the unique advantages of India for work in this field - wide range of latitudes from equator in the South to 25 degrees North in Kashmir within the boundaries of a single country; mountain stations in the south and north and the deepest mines in the world. Millikan had come all the way from U.S. to do experiments at several stations in India in the mid 30's.

With a uniquely designed GM telescope, which Bhabha built with the help of S.V.C. Iya, the penetrating particle intensities were measured at altitudes of 5000, 10,000, 15,000, 20,000, 25,000 and 30,000 ft, using a B-29 bomber Air Craft belonging to the U.S. Air Force. These constituted the first measurements at such high altitudes in an equatorial latitude. Comparison with the measurements of Schein, Jesse and Wollan in the U.S.A., established that no marked increase of intensity occurred between 3.3 degrees north and 52 degrees north even at an altitude of 30,000 ft., in contrast to the total intensity which exhibited very pronounced latitude effect at such altitudes.

At the Indian Institute of Science, Bhabha also got constructed a 12" diameter cloud chamber identical to the one operating in Manchester. R.L. Sengupta, who had worked in Blackett's Laboratory helped Bhabha in the design and construction of this chamber, which was used by M.S. Sinha to study the scattering characteristics of mesons. Vikram Sarabhai set up a telescope to study the time variation of cosmic ray intensity.

Bhabha and Cosmic Ray Research at the Tata Institute of Fundamental Research

While at the Indian Institute of Science, Bhabha recognised the need for setting up in the country an institute solely devoted to the pursuit of fundamental research especially in the area of nuclear science that was emerging as a virgin area of fundamental science. The developments in the field of cosmic ray studies and in the area of nuclear physics with accelerators had convinced Bhabha that the future lay in these areas. With financial support from the Sir Dorabji Tata Trust and the Government of Maharashtra, Bhabha established the Tata Institute of Fundamental Research in Bombay in June 1945. The TIFR became an aided institution under the Department of Atomic Energy later and was recognised as the National Centre for Nuclear Science and Mathematics by the Government of India. Bhabha himself used to say that TIFR was the "Cradle of the Atomic Energy Programme" of the country. Bhabha was the Director of TIFR from 1945 to January 1966 - till his untimely death in a tragic air crash on the Alps.

The TIFR naturally started with a major experimental programme in cosmic rays, taking cognisance of the fact that cosmic ray research had entered its second phase the world over. The Pi-meson as the parent of the Mu-meson was discovered in 1947 by Powell and his collaborators at the university of Bristol exposing the newly developed high sensitivity nuclear emulsions in the Jangfraujoch mountains in Switzerland. Rochester and Butler discovered the same year the V^{o} particles, which were later identified as the K-mesons and Hyperons through nuclear emulsion experiments by several groups. Bradt and Peters discovered around the same time the presence of α-particles and other stripped heavy nuclei in the primary cosmic radiation which consisted predominantly of protons. Also, most importantly, the act of meson production had been caught both in nuclear emulsions and in multiplate cloud chambers.

With these developments, the new directions of cosmic ray research had become clear. To enter the international arena in this field, the emphasis had to be on (i) the investigations on the primary component - spectrum, composition, anisotropy of arrival directions; relative proportions of rare nuclei, electrons, gamma-rays (ii) the detailed study of the characteristics of nuclear collisions of the primaries as well as of the secondaries produced in these collisions (iii) the studies on the penetrating components - muons and neutrinos in deep underground installations (iv) studies on the Extensive Air Showers initiated in the atmosphere through the nuclear and electromagnetic cascades by the primaries (v) Study of the radio isotopes produced by cosmic rays (vi) time variation studies on cosmic ray intensity and correlations with solar activity.

These multidimensional studies to be carried out in a variety of locations with specially designed detector systems, required the development inhouse of a variety of technologies - to name a few - Plastic Balloon Fabrication Technology, fabrication of GM counters, plastic scintillators, multiplate cloud chambers, pulsed electronic circuits and even a digital computer. Thanks to the organisational genius of Bhabha, all this was done in a record time in TIFR itself. The Indian industry was very backward in the 40's and 50's and import was just not thinkable because of shortage of foreign exchange and the enormous delays of transportation. The cosmic ray programme did get a fillip in the 50's by Bernard Peters, the co-discover of heavy primaries and M.G. K. Menon who worked for 8 years in Powell's Laboratory, joining the Nuclear Emulsion

Group of TIFR.

At the International Conference on Cosmic Rays held at Bagneres in 1953, TIFR made its first impact by presenting very significant results on K-Mesons and Hyperons obtained from the analysis of emulsion stacks exposed at Hyderabad. The emulsion group kept a high profile of original contributions in the field of high energy interaction studies, the relative abundances of Li, Be and B in the primaries, Hyperfregments and on the spectrum of primary electrons. The deep underground experiments in the Kolar Gold Fields initiated at the instance of Dr. Bhabha as early as 1950, and which continued for more than four decades, till 1994, was another line of activity in which pioneering contributions were made - most accurate mu-meson intensity and angular distribution measurements upto very high energies, detection of neutrino induced interactions with a visual detector, limits on the lifetime of protons etc. These involved very large scale installations and also international collaborations. Extensive Air Shower Array with a variety of detectors for different components, - scintillators, Cerenkov Counters, Total Absorption Spectrometer, Multiplate Cloud Chamber started operating in the late 50's in the mountain station at Ooty - the time structure measurements of hadrons with the Total Absorption Spectrometer, led to the first recognition of increased cross section for the production of nucleons and antinucleons at high energies. Dr. Bhabha, when he visited the Ooty Laboratory in 1964, was thrilled to see the world's largest multiplate cloud chamber operating there. This cloud chamber gave unique information on the highest energy jets produced by the incidence of several parallel hadrons. At the Kolar Gold Fields, a second air shower array was set up at the surface of the mines with large area detectors at several depths underground that recorded the associated very high energy muons. This set up gave very valuable information on the composition of the primaries in the crucial knee region $10^{14} - 10^{16}$ ev.

In a short article like this it is difficult to do full justice and bring out the full flavour and ramification of all the work in cosmic rays that got initiated at the instance of Dr. Bhabha. Dr. Bhabha's was a multidimensional, many splendoured personality that influenced not only Cosmic ray research, but many other fields too. But cosmic rays were very dear to him, all through his life, may be because his very first paper was on Cosmic Rays.

Even 86 years after the discovery of Cosmic Rays, 50 years after entering the second phase, despite colossal efforts by groups all over the world, not a single source of cosmic rays of high energy (> 20 Gev) has been identified even though it is firmly established that the spectrum extends beyond 10^{20} ev. The mechanism by which particles are accelerated to such high energies is also not known. The high rotating magnetic field environments of the neutron stars in Pulsars in the galaxy and the extraglactic Active Galactic Nuclei with suspected giant blackholes in their centre are thought to be the strongest candidates. Gigantic Multiplex installations are coming up to settle this question. What other exotic particles are there among the primaries and what new particles are produced in super high energy collisions are other aspects which are receiving special attention in the design of next generation cosmic ray experiments.

Resonance

July 1998 Volume 3 Number 7

journal of science education

Homi Bhabha and Cosmic Ray Research in India ❖

Enzyme Kinetics? Elementary my dear ... ❖

Game Theory ❖ Machine Translation ❖

Teaching the Limit Concept

Resonance is a monthly journal of science education published by Indian Academy of Sciences, Bangalore, India.

Editorial Office: Indian Academy of Sciences, C V Raman Avenue, PB No. 8005, Bangalore 560 080, India.
Tel:+91 (80) 3342310 / 3342546, Fax:+91 (80) 3346094, email:resonanc@ias.ernet.in

Editorial

N Mukunda, Chief Editor

A recently published book *Portraits of Discovery* by astronomer and science writer George Greenstein, describing the lives and achievements of ten scientists from different parts of the world over the past century or so, refers to Homi Jehangir Bhabha as "A Gentleman of the Old School". It is often illuminating to see what others, far removed from us, have to say about us and our problems. Presuming that there is a basic sympathy in outlook, we can expect some objectivity in their assessments. In his Prologue Greenstein says:

"Most of Homi Bhabha's work was concerned with the development of science in India, a nation beset by problems so overwhelming as to make the practice of science incomparably more difficult than here."

There seems to be a hint that even in the USA the pursuit of science faces difficulties! Towards the end of his account Greenstein remarks:

"In thinking of this man, the image perpetually rises to my mind of one of those great, larger-than-life figures of the Renaissance …".

Bhabha was an aristocrat in many ways: by birth, in his tastes, and in his way of life. There also seems to have been a distance between him and those with whom he otherwise worked closely. But all this seems to have been necessary for him to have had visions on a truly grand scale, and to have translated them into reality. Quoting Greenstein again:

"There is not the slightest doubt in my mind that Bhabha would never have achieved what he achieved had it not been for his aristocratic background and personal connections."

A great deal has been written on the contrasting attitudes of Homi Bhabha and, say, Megh Nad Saha on what needed to be

"Most of Homi Bhabha's work was concerned with the development of science in India, a nation beset by problems so overwhelming as to make the practice of science incomparably more difficult than here."

— George Greenstein

"Art, music, poetry, and everything else that I do have this one purpose – increasing the intensity of my consciousness and life".
— Homi Jehangir Bhabha

done to revitalise Indian science. Here we want to tell our readers something about Bhabha's achievements in physics and other creative pursuits. G Venkataraman, the author of *Homi Bhabha and his Magnificent Obsession* published not long ago, gives a brief life sketch: Bhabha's educational career, the years in Cambridge and in Europe devoted to physics, and then the Indian period. At Cambridge Bhabha was a student of R H Fowler for his PhD (So were, incidentally, Paul Dirac and Subrahmanyan Chandrasekhar). Then comes the period he spent at the Indian Institute of Science, Bangalore, as a result of circumstances connected with World War II. Through extracts from his letters to his parents and others, on various pages of this issue, we see an expression of his passion for physics; and later of his realisation that he saw a mission for himself – the creation of conditions in India where world class science could be pursued.

B V Sreekantan (a student of Bhabha) describes the origins of cosmic ray physics, Bhabha's deep involvement in this field, and how after the decision to settle in India he created and led a major experimental and theoretical effort in this area. Bhabha's best remembered contributions to physics are his analysis of electron-positron (Bhabha) scattering, and the cascade theory of cosmic ray showers.

In the Reflections section we reproduce Bhabha's scholarly presidential address to the 1951 session of the Indian Science Congress. Considering the occasion, the range of ideas covered is remarkable: the meaning and philosophy of natural law, the importance of quantification, the developments of relativity and quantum mechanics, and elementary particle physics at that time. As fruits of his artistic talent, we present a couple of his paintings. And what better way to conclude than by quoting Bhabha himself:

> *"Art, music, poetry, and everything else that I do have this one purpose – increasing the intensity of my consciousness and life".*

Homi Bhabha – A Profile

These days, young people dream of going abroad even before they have completed their studies, often for settling there permanently. Nearly sixty years ago, a young man made the journey in the reverse direction. After spending thirteen years in Cambridge, at that time the Mecca of Physics, Homi Bhabha, then aged twenty nine, came back to India not only to settle down permanently but to change her destiny as well.

Homi Jehangir Bhabha was born on 30 October, 1909 in Mumbai (then Bombay). The house he was born in was later destined to be the cradle of India's Nuclear Energy Programme! Young Homi was educated at the Cathedral and John Connon High School. Absolutely brilliant in studies, he became a minor celebrity. Homi was a voracious reader and his father's wonderful collection helped him to greatly broaden his outlook. In addition, he was also keenly interested in art as well as music (particularly western).

After passing the Senior Cambridge Examination and studying for a couple of years in the Royal Institute of Science, Mumbai, Bhabha went to Cambridge for higher studies. His father wanted Homi to specialise in mechanical engineering so that he could, on return, enter the corporate world of the Tata group of industries and rise to a high position there. But things did not work out that way. Bhabha found that right then, physics was going through a major revolution, a good bit of the action being in Cambridge itself. So he wrote to his father:

> *I seriously say to you that business or job as an engineer is not the thing for me. It is totally foreign to my nature and radically opposed to my temperament and opinions. Physics is my line ... I am burning with a desire to do physics. I will and must do it sometime. It is my only ambition.*

The father was understanding and allowed Homi to study for the Mathematical Tripos, after completing the Mechanical Tripos. In 1932, Bhabha won the Rouse Ball Travelling Fellowship which enabled him to work with Pauli in Zurich and Fermi in Rome. Later, the Isaac Newton Studentship allowed him to spend some time in Niels Bohr's Institute in Copenhagen. In between, Bhabha completed his PhD thesis under the supervision of R H Fowler, who was also the supervisor for S Chandrasekhar.

In Cambridge, Bhabha discovered what is now referred to as *Bhabha scattering*, a phenomenon whose existence has been confirmed experimentally. In addition, he developed, in collaboration with Walter Heitler, a theory for cosmic-ray showers, known as the *cascade theory*. Both these contributions made Bhabha quite well known in physics circles.

In early 1939, Bhabha came back to India for what was supposed to be a brief holiday. Meanwhile, the second World War broke out, and the holiday turned into a permanent stay. War severely disrupted the scientific scene in Europe, and it was clear that Bhabha would have to look for a job in India. Thanks to his reputation, he received a few offers from some Universities but in the end, he joined the Indian Institute of Science, Bangalore. Here, with a small grant from the Sir Dorab Tata Trust, he started working on cosmic rays.

In the Bangalore period, Bhabha concentrated mainly on theory and discovered what is known as the *Bhabha equation.* Apart from this, he briefly collaborated with Harish-Chandra, later to win fame as a mathematician. Bhabha also tried his hand with experiments, building Geiger-counter telescopes and flying them in air force planes, to study cosmic ray behaviour at high altitudes.

By 1944 it became clear that the war, at least in Europe, was drawing to a close. Bhabha was in two minds about what he should do. Should he go back to the West, which offered so many opportunities, or should he stay? He wrote to his friend J R D Tata seeking his advice, adding that he was ready to continue in India as "it is one's duty to stay in one's own country and build up schools comparable in other lands." J R D encouraged Bhabha to approach the Sir Dorab Tata Trust. Bhabha promptly did so, and in March 1944 he wrote to the Trust seeking grants and promising

> *to build up in the course of time a School*
> *of Physics comparable with the best anywhere.*

Prophetically, he also declared that when the time came, the School would provide the experts needed for exploiting nuclear energy, and that India would not have to look for such experts overseas.

Events now moved rapidly. With a small grant of less than Rs 1 lakh per year (of which the Tata Trust's contribution was Rs 45,000/-) Bhabha founded the TIFR on June 1, 1945. It started functioning first in Bangalore but by December, Bhabha had TIFR shifted to Mumbai, locating it in the very house he was born in!

In the beginning TIFR concentrated exclusively on cosmic rays and mathematics, but as bright young people came forward to join it, it rapidly expanded in size as well as scope. Meanwhile the country became independent, and Bhabha, on account of his closeness to Nehru, was given the task of steering the country's nuclear energy programme. Bhabha's enthusiasm was infectious, and it was like a breath of fresh air in a country notorious for its negative and bureaucratic thinking (which, alas, persists). Both TIFR and the Indian Atomic Energy Programme blossomed in a few short years to proportions unimaginable and beyond all expectations.

Bhabha was a thorough-bred theoretical physicist, deeply influenced by Dirac. Yet, when it came to matters of technology, he was second to none. He also laid the seeds for our very successful space programme, which later Vikram Sarabhai and Satish Dhawan nursed with loving care. After the Chinese attack on India in 1962, Bhabha realised our backwardness in electronics and worked hard to prepare a masterplan to help the country leapfrog in this vital area. Unfortunately, he died before the report he had prepared (called the Bhabha Committee Report) could be submitted. Later, like all reports, this one too collected dust, and we never made the big jump in electronics we could have.

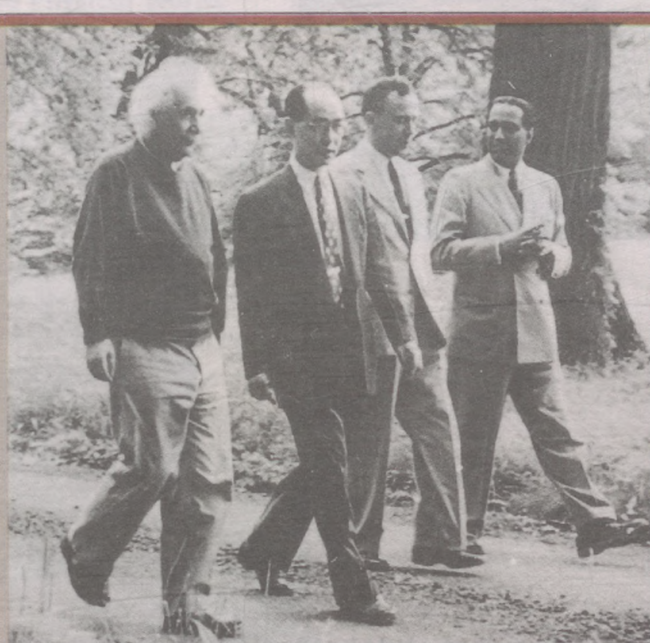
Bhabha was riding high, all the time reaching new pinnacles, but death came suddenly. In January 1966, Bhabha was on his way to Vienna to attend a meeting. As his Air India plane descended to land in Geneva, it crashed into Mont Blanc. All the passengers including Bhabha perished. His body was never found. For the nation, it was an irreparable loss.

Bhabha was not merely a gifted scientist, and an able technocrat; he was much more. He was an artist, and a connoisseur of all the good things in life – art, music, literature, architecture, landscaping, gardening, ... Many hailed him as a modern Leonardo while JRD referred to him as an authentic genius. He represented the best in both science as well as culture, often regarded as unbridgeable. Expressing the sorrow of the nation, Indira Gandhi said of him:

> *He was a scientist of great originality. He was an artist endowed with unusual sensitivity. His interest in music was as serious as it was deep. The flower beds, the landscaping, the architecture of buildings in Trombay, all bear witness to Homi Bhabha's perception of colour, form and design. India will long cherish Homi Bhabha's memory, for he was deeply involved in her destiny and in the process of changing the texture and quality of her society.*

Way back in 1928, Bhabha told his father: "Who says we can't do science in India?" Not only did he convincingly demonstrate later that he could, but, more important, he made it possible for hundreds if not thousands of others also to do likewise.

G Venkataraman
Vice-Chancellor
Sri Sathya Sai Institute of Higher Learning, Anantapur, India



Albert Einstein, Hideki Yukawa, John Archibald Wheeler and Homi Jehangir Bhabha in Princeton around 1950.

Resonance journal of science education

July 1998 Volume 3 Number 7

Portrait of a lady by Bhabha

DEPARTMENTS

Front Cover

An abstract painting by Homi J Bhabha

Back Cover

Bust of Dr Homi J Bhabha (1909–1966)
at the Entrance Foyer, TIFR, Mumbai
Presented by J R D Tata on behalf of
M/s Tata Sons Limited on 24-1-1967
Sculptor: B Vithal

Molecules to Materials

3. Molecular Magnetic Materials

T P Radhakrishnan

T P Radhakrishnan is in the School of Chemistry, University of Hyderabad, Hyderabad 500 046.

Molecular materials represent an area of fruitful interaction between synthetic chemistry, solid state physics and materials science. This article discusses the development of magnetic materials based on metal complexes, organometallic compounds and organic radicals.

Previous articles of this series were on:

1. Design and fabrication of functional molecular solids, Vol.3, April 1998.
2. Liquid crystals and molecular conductors, Vol.3, May 1998.

Please Note in part 2.

1) Page 17 last line was inadvertently omitted, the same is given below: '... 6.3×10^5 Scm^{-1} and 2.2×106 Scm^{-1}. However, at lower ...'

2) Figure 5: The structure on the right is TTF$^{+\cdot}$ (and not TTF^{+-}).

3) Figure 6: Real 'unpaired e$^-$' instead of unpaired $^-$e.

Introduction

In Part 1 of this series of articles the historical background of the development of molecular materials was presented. The subtle interactions exploited in their design and the various methods of fabrication were reviewed. Part 2 provided an overview of liquid crystals and molecular semiconductors, conductors and superconductors. In the present article we focus attention on molecular magnetic materials.

We outline the early development of magnetic materials based on coordination polymers and molecular systems in which metal ions serve as the source of the magnetic moment. The intense research efforts that followed to develop magnetic materials based on organic radicals are reviewed.

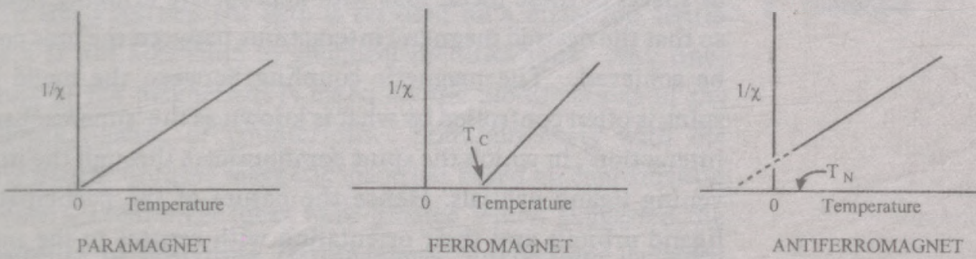
Molecule-based Magnetic Materials

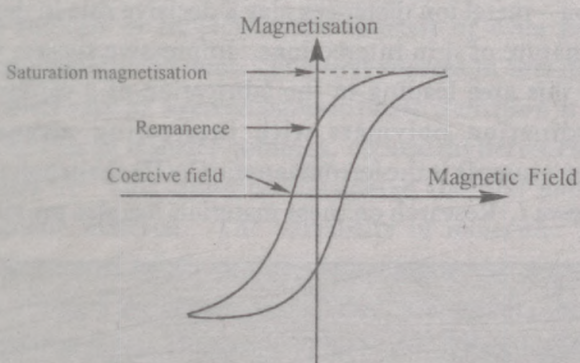
For centuries after the discovery of the naturally occurring magnet, Fe_3O_4, most of the magnets that were fabricated and used were based on the compounds and alloys of elements such as iron, cobalt, nickel and gadolinium which are themselves ferromagnetic in their bulk state (*Box 1*). Therefore the idea of designing and fabricating a ferromagnetic solid starting with simple paramagnetic units and controlling its magnetism using synthetic manipulations, emerged as a fascinating problem. Naturally, the initial attempts in this direction were based on paramagnetic

The Circulation Department
Indian Academy of Sciences
P. B. No. 8005, C V Raman Avenue
Bangalore 560 080

Box 1. Paramagnetism, Ferromagnetism, Antiferromagnetism and Ferrimagnetism

Atoms, ions, molecules and solids having unpaired electrons when brought under the influence of a magnetic field, develop a net magnetisation (alignment of the spins) and are attracted into the field. Such materials are described as paramagnetic. In the absence of unpaired electrons, paramagnetism can still show up if the nucleus has a spin; however, nuclear paramagnetism is usually much weaker than electron paramagnetism. Systems having only paired electrons are described as diamagnetic and these are repelled from a magnetic field. Most organic molecules are diamagnetic. Metal ions such as Mn^{2+} and Cu^{2+} and their complexes, organic free and ion radicals and molecules such as O_2 and NO are good examples of paramagnetic systems. Paramagnetism is characterised by the magnetic susceptibility, χ, the variation of the magnetisation with respect to the applied magnetic field. The paramagnetic susceptibility varies inversely with the temperature and this behaviour is known as the *Curie law*.





HYSTERESIS IN A FERROMAGNET

In some materials such as iron, cobalt and nickel the spins can remain aligned parallel, even in the absence of an external magnetic field. This is possible due to an internal field arising as a result of a cooperative interaction between the spins. Such materials having a spontaneous magnetisation are referred to as ferromagnets. It can also happen that the internal field causes each spin to align antiparallel with respect to its nearest neighbours. Such a phenomenon is called antiferromagnetism. Famous examples of antiferromagnets are metal compounds such as MnO, MnF_2 and NiO. Ferrimagnetism is a special case of antiferromagnetism, where the neighbouring spins which are aligned antiparallel are of unequal magnitude so that a total cancellation of the magnetic moments does not occur. The best example of a ferrimagnet is the mineral 'magnetite' (Fe_3O_4). At certain characteristic temperatures, ferro, antiferro and ferrimagnets undergo phase transition to the paramagnetic state. This temperature in the case of a ferromagnet is referred to as the *Curie temperature*. In the paramagnetic regime of ferromagnetic materials, the susceptibility rise with lowering temperature is stronger than in the case of simple Curie paramagnets. In the paramagnetic regime of antiferromagnets (*i.e.* above its phase transition temperature, namely the Néel temperature, T_N), the susceptibility increase on lowering temperature is weaker than in paramagnets. Several other kinds of complex magnetic orderings are possible in solids.

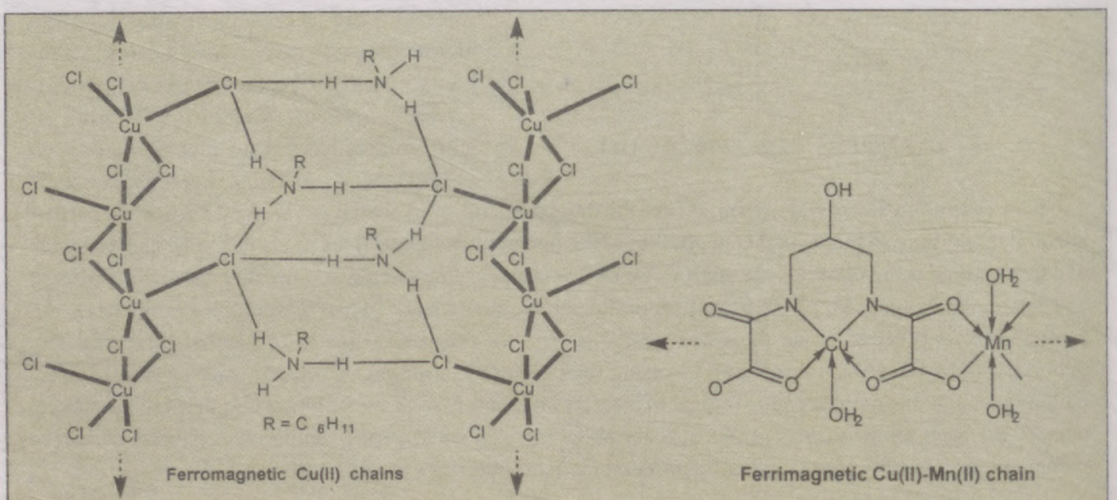
Continued...

Continued from page 9.

When a ferromagnet is subjected to an increasing magnetic field, the magnetisation increases and reaches a constant value known as the saturation magnetisation. If the applied field is then reversed, the magnetisation decreases but does not become zero when the applied field is reduced to zero. The magnetisation at zero field is called *remanence*. The reversed magnetic field required to bring the magnetisation back to zero is known as the *coercive field*. This nonreversible behaviour of the magnetisation against the applied magnetic field is termed *hysteresis* and is an important characterisation of a ferromagnet. The magnitude of the spontaneous magnetisation, remanence and coercive field determine the type of application that a ferromagnet can be put to.

metal ions as the spin centres. The basic idea is to build chains or sheets of these metal ions with appropriate bridging ligands so that the desired magnetic interactions between the ions could be achieved. The magnetic coupling between the metal ion spins is often controlled by what is known as the 'superexchange interaction', in which the spins communicate through the intervening ligand orbitals. Hence the nature of the participating ligand orbitals and their orientation with respect to the metal orbitals are crucial. The metal ion – ligand – metal ion angles and metal ion – metal ion distances play a decisive role in determining the nature of spin interactions. Impressive success was achieved in this area leading to the fabrication of a variety of metal coordination polymers with interesting magnetic properties; an example is the ferromagnetic Cu(II) chain polymer shown in *Figure 1*. Research on these materials has also provided

Figure 1. Examples of ferromagnetic and ferrimagnetic coordination polymers.



a sound understanding of the fundamental principles of magnetic interactions in low-dimensional lattices.

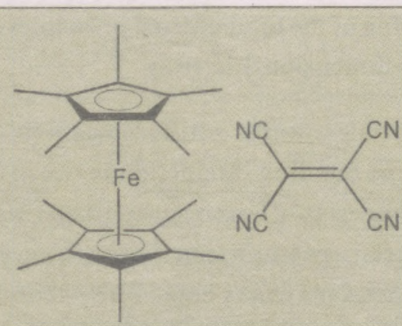
The dominant tendency towards antiferromagnetic coupling of spins arising from the bonding interaction between orbitals is a serious problem in the design of ferromagnetic materials. However, this inherent antiferromagnetic interaction has been exploited by designing ferrimagnetic polymers based on systems such as the mixed Mn-Cu complexes (*Figure 1*) wherein the Mn^{2+} and Cu^{2+} ions bear respectively 5/2 and 1/2 spins.

The dominant tendency towards antiferromagnetic coupling of spins arising from the bonding interaction between orbitals is a serious problem in the design of ferromagnetic materials.

In all these systems the spin is resident on a metal ion centre which is not amenable to chemical modifications. Any fine-tuning of the magnetism is based on the modifications of the bridging ligands or counterions. Simultaneously with the development of these magnetic materials based on coordination polymers, theoretical ideas were mooted which suggested the possibility of designing ferromagnetic interactions between molecular spin systems; the basic spin units would be organic free radicals or molecular ion radicals or even paramagnetic organometallic systems. The first success in this direction was the discovery of a ferromagnetic phase transition at about 4.8 K in the charge transfer complex, decamethylferrocenium-TCNE (*Figure 2*). This could be considered the first molecule-based magnetic material. The versatility of molecular materials is demonstrated by the profound changes in the magnetic properties effected by alterations of the components in this charge transfer complex (*Figure 2*). The latest in this class of materials is the TCNE complex of vanadium containing molecules of the solvent, dichloromethane ($V(TCNE)_x.yCH_2Cl_2$). This compound is found to be ferromagnetic at ambient temperatures; however, it is handicapped by extreme chemical reactivity and the consequent need to preserve it under inert atmosphere. The origin of ferromagnetism in these materials has been a subject of considerable debate. The contribution of high spin charge transfer excited states was considered important in the mechanisms proposed initially. It has been argued later that the direct interaction of spin densities between the radical

The first success in the design of molecule based magnets was the discovery of a ferromagnetic phase transition at about 4.8 K in the charge transfer complex, decamethyl-ferrocenium - TCNE.

Figure 2. Decamethylferrocene-TCNE complex; magnetic properties of charge transfer complexes of metallocenes with organic π-acceptors.



Decamethylferrocenium-TCNE

$[Fe^{III}(C_5Me_5)_2]^{+\cdot}[TCNE]^{-\cdot}$

$[Fe^{III}(C_5Me_5)_2]^{+\cdot}[C_3(CN)_5]^{-}$	paramagnetism
$[Ni^{III}(C_5Me_5)_2]^{+\cdot}[TCNE]^{-\cdot}$	antiferromagnetism
$[Fe^{III}(C_5Me_5)_2]^{+\cdot}[TCNE]^{-\cdot}$	ferromagnetism
$[Cr^{III}(C_5Me_5)_2]^{+\cdot}[TCNE]^{-\cdot}$	ferrimagnetism
$[Fe^{III}(C_5Me_5)_2]^{+\cdot}[TCNQ]^{-\cdot}$	metamagnetism

components is more significant. We discuss the latter mechanism in the following section.

Organic Magnetic Materials

Fabrication of a completely metal-free ferromagnet continued to be a challenge for materials chemists. Since purely organic radicals are not difficult to come by (*Figure 3*), the focus of the effort is to attain a ferromagnetic coupling of these spins in the solid state. Unlike in metal-based systems where the unpaired electron spin is largely confined to the metal ion centre, in molecular spin systems the unpaired electron is delocalised over the various atoms that constitute the molecule (compare with the discussion of charge distribution in Part 1 of this series). The contribution of each atomic site to the total spin is called the spin density at that site; these spin densities can be positive

Fabrication of a completely metal-free ferromagnet continued to be a challenge for materials chemists.

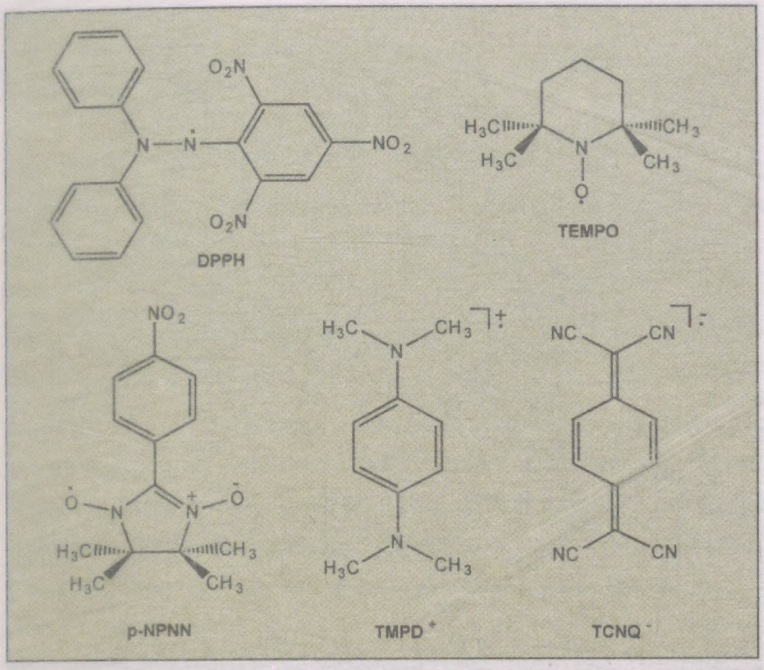


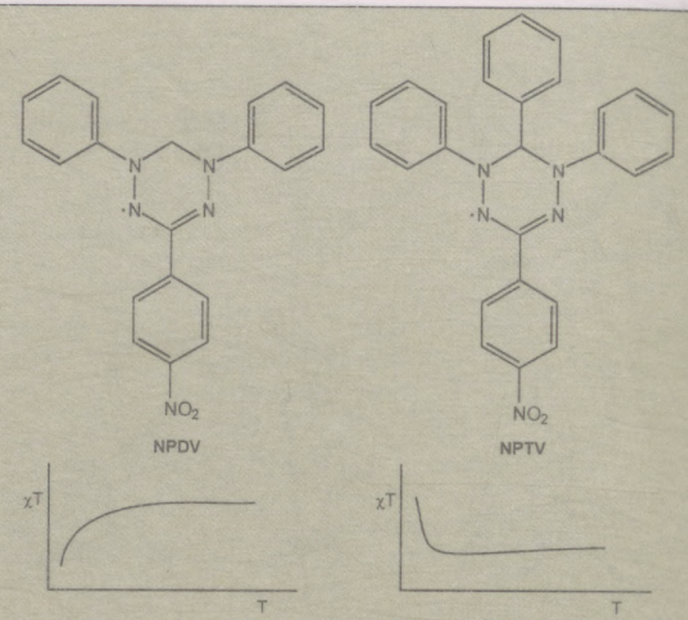
Figure 3. Examples of stable organic radicals, neutral and ionic.

or negative. As noted earlier, bonding tendencies between orbitals lead to preferential antiferromagnetic coupling of spins. One of the ideas proposed to obtain ferromagnetic spin coupling between molecular spin systems was to achieve a packing of the molecules in such a way that the regions of positive spin densities on one molecule would be in close contact with regions of negative spin densities on the neighbour and *vice versa*. This would make the products of spin densities on one radical with the spin densities on the second radical predominantly negative, which in turn coupled with the inherent negative exchange interactions lead to an effectively positive or ferromagnetic spin coupling.

One proposal to obtain ferromagnetic coupling between molecular spin systems is to 'engineer' the packing of the molecules such that regions of positive spin densities of one molecule are close to regions of negative spin densities of its neighbours and vice versa.

The following is a fine experimental demonstration of how molecular design influences the nature of spin coupling between neighbouring radicals as dictated by the above principle. *p*-Nitrophenyldiphenylverdazyl (NPDV in *Figure 4*) is a stable radical that shows antiferromagnetic spin interactions at low temperatures as seen from the dip of the χT value. When an extra phenyl group is substituted on the verdazyl moiety to make NPTV, the low temperature spin interactions do a *volte-face* and become ferromagnetic. This interesting phenomenon

Figure 4. Ferro and antiferromagnetic interactions in verdazyl based radicals.



Schematic plots of χT (magnetic susceptibility × temperature) vs. temperature.

may be explained on the basis of the spin density interaction model. In the crystals of NPDV, the adjacent radicals have regions of the same spin densities in close proximity. The extra phenyl group in NPTV creates a steric hindrance that causes one molecule to slip with respect to its neighbour so that regions of opposite spin densities on the neighbouring radicals come close together. Thus the mode of spin coupling is antiferromagnetic in NPDV and ferromagnetic in NPTV. Though ferromagnetic spin interaction is obtained in NPTV, this crystal does not undergo a phase transition to a ferromagnetically ordered state. The idea of spin density interaction was successfully implemented in crystals of the stable radical, *p*-nitrophenylnitronylnitroxide (*p*-NPNN in *Figure 3*) which undergoes phase transition to a ferromagnetically ordered state below 1K. Though the Curie temperature is extremely low, this material provided the first example of a ferromagnet based on a purely organic molecule containing only C, H, N and O atoms. Several related systems based on the nitronylnitroxide radical are found to become ferromagnetic at similar low temperatures.

Crystals of *p*-NPNN below 1K provided the first example of a ferromagnet based on a purely organic molecule containing only C, H, N and O atoms.

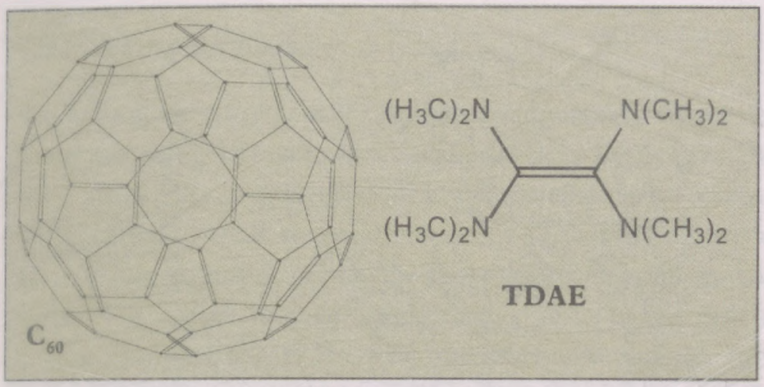
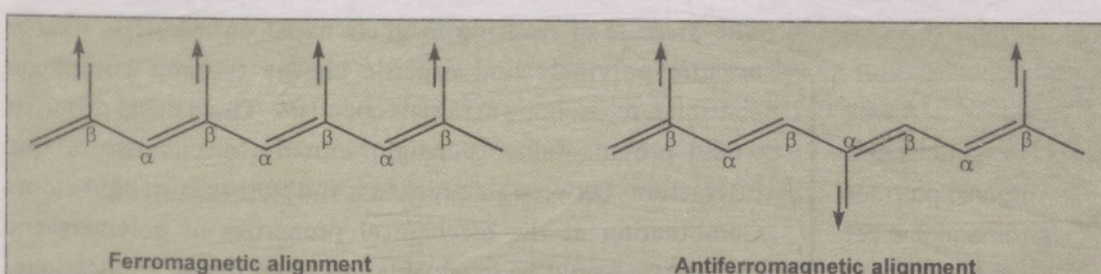


Figure 5. Buckminsterfullerene (C_{60}) and tetrakis (dimethylamino)ethylene (TDAE).

Another interesting case of an organic ferromagnet is based on, by now familiar, buckminsterfullerene (recall the C_{60} based superconductors mentioned in Part 2 of this series). The charge transfer complex between the powerful electron donor TDAE and C_{60} (*Figure 5*) was found to undergo a phase transition to a ferromagnetic state at about 16K. The resulting magnetic material is found to have small coercive fields and is termed a soft ferromagnet. Like the *K*-doped fullerene superconductors, this material also is plagued by extreme chemical reactivity.

There have been extensive efforts to realise experimentally, ferromagnetic spin interactions in organic oligo and polyradicals. A simple connectivity principle can be used to visualise the magnetic interactions. If radical sites occur at locations separated by an odd number of π-electrons on a conjugated polymer, the resulting spin polarisation along the conjugation pathway should lead to a ferromagnetic alignment of the spins at the radical sites; when the sites are separated by an even number of electrons, the magnetic coupling will be antiferromagnetic (*Figure 6*). Basic entropy considerations show that long range ordering is

Figure 6. Representation of spin polarisation leading to ferro and antiferromagnetic spin coupling in polyradical fragments.



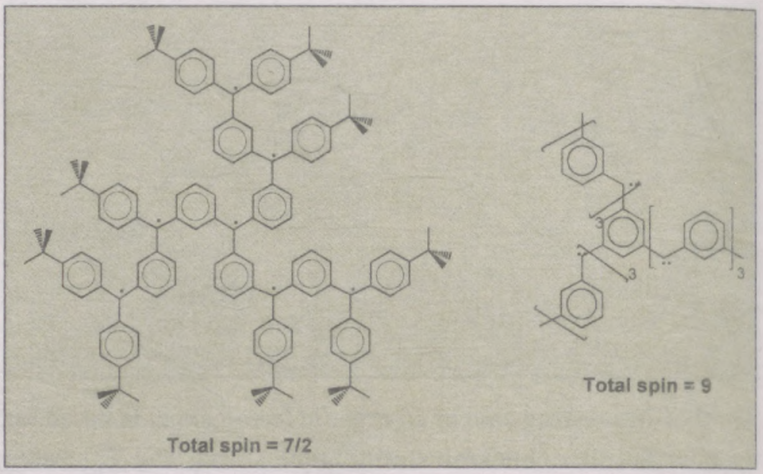


Figure 7. Examples of high spin organic oligoradicals. Note that the structure on the right is made of carbene units, each with two unpaired electrons, leading to a grand total of 18 parallel spins.

impossible at finite temperatures in one-dimensional systems; taking cognisance of this, ferromagnetic materials may be designed based on π-conjugated two-dimensional networks with appropriately placed radical sites. These ideas have been discussed at the theoretical level in several different ways and extensive experimental work has been carried out. Even tentative claims of the synthesis of polymeric ferromagnets have been made and disproved later. Success has been limited to achieving small spin clusters based on some carefully designed and synthesised compounds; some interesting examples of high spin oligoradicals reported in the literature are provided in *Figure 7*. A well characterised organic polymer ferromagnet is yet to be realised. One of the major problems with organic polyradicals is the high chemical reactivity that results in extensive loss of the spin centres. Investigation of composite systems containing paramagnetic metal ion spin sites and stable organic radicals like nitrosyl molecules is becoming increasingly popular.

A well characterised organic polymer ferromagnet is yet to be realised.

The promise of realising magnets based on materials such as organic polymers and organic charge transfer complexes continues to fascinate materials chemists. These novel materials would provide valuable insight into the mechanism of spin interactions between electrons in *s* and *p* orbitals on light atoms. Combination of the mechanical properties of polymers and magnetism would be invaluable in several device applications.

Concluding Remarks

We have made a brief survey of the development of molecule-based magnetic materials . Once again the focus has been on the basic philosophy of optimisations at the molecular level leading to the control of bulk material properties. The unifying theme of how the assembly of the molecular units determines the material attributes is underscored throughout. In the case of magnetism, this has to do with the interaction of the spin densities of adjacent radicals. In the next part of the series, we will consider potential non-linear optical applications of molecules and polymeric systems.

Combination of the mechanical properties of polymers and magnetism would be invaluable in several device applications.

Suggested Reading

O Kahn. *Molecular Magnetism*. VCH. New York, 1993.

Address for Correspondence
T P Radhakrishnan
School of Chemistry
University of Hyderabad
Hyderabad 500 046, India

I seriously say to you that business or job as an engineer is not the thing for me. It is totally foreign to my nature and radically opposed to my temperament and opinions. Physics is my line. I know I shall do great things here. For, each man can do best and excel in only that thing of which he is passionately fond, in which he believes, as I do, that he has the ability to do it, that he is in fact born and destined to do it. My success will not depend on what A or B thinks of me. My success will be what I make of my work. Besides, India is not a land where science cannot be carried on ... I am burning with a desire to do physics. I will and must do it some time. It is my only ambition. I have no desire to be a 'successful' man or the head of a big firm. There are intelligent people who like that and let them do it. I hear you saying 'But you are not Socrates or Einstein'. No – and that is what Berlioz's father said to Berlioz. Berlioz who is now accepted as one of the world's greatest geniuses and France's greatest musician. How can anybody else know at what time what one will do, if there is nothing to show. ... It is no use saying to Beethoven 'You must be a scientist for it is a great thing', when he did not care two hoots for science; or to Socrates 'Be an engineer; it is the work of an intelligent man'. It is not in the nature of things. I therefore earnestly implore you to let me do physics.

Homi J Bhabha in a letter to his father, 1928.

Homi Bhabha and Cosmic Ray Research in India

B V Sreekantan

B V Sreekantan is currently the Sir S Radhakrishnan Visiting Professor at the National Institute of Advanced Studies, Bangalore. He was Director of Tata Institute of Fundamental Research, Mumbai during the period 75-87. He obtained his PhD from Bombay University in 1954 for his work on 'underground cosmic rays' under the guidance of Homi Bhabha.

Cosmic rays are very high energy particles arriving from the depths of space and incident on the earth's atmosphere at all places and at all times. The energy of these particles extends over 12 decades from around 10^9 ev to 10^{21} ev and mercifully for the survival of life, the intensity falls by atleast 22 decades from about 100 particles/cm^2/s to 1 particle/1000 km^2 year. Cosmic ray research led to the discovery of many of the fundamental particles of nature in the 30's, 40's and 50's of this century and ushered in the era of 'elementary particle physics' at man-made accelerators. Even 86 years after the discovery, the sources of these particles and the mechanism of acceleration continue to remain a mystery.

Homi Bhabha who became famous for his 'cascade theory of the electron' in the 30's did pioneering theoretical and experimental research in this field during his post doctoral fellowship in Cambridge and later at the Indian Institute of Science in Bangalore. The Tata Institute of Fundamental Research, which he founded in 1945, became under his leadership, a major centre of cosmic ray research covering practically all aspects of the radiation and continues to be active in this field.

Homi Bhabha is well known among the scientists of India and the public at large as the 'Father of India's Atomic Energy Programme'. What is not so well known however, particularly among the younger generation is the fact that he made outstanding contributions in the field of theoretical physics and played a very important role in initiating and fostering cosmic ray research in India. Thanks to the unique start given by him in the 40's, cosmic ray research in India grew into one of the largest activities in the world covering all aspects of the radiation in the 50's and 60's. In recognition of this, the Cosmic Ray Commission of the IUPAP held its 7th International Conference on Cosmic Rays, one of the earliest in the series, at Jaipur in India in 1963. It is most interesting and instructive to know how all this was achieved and what exactly motivated a young man in his thirties to initiate work in a highly sophisticated and highly technology dependent field so early in India. Let us look at the background of Bhabha and the times in which he began his research.

In 1934, Bhabha was awarded the Isaac Newton Studentship at Cambridge which enabled him to complete his PhD under R H Fowler.

Homi Bhabha at Cambridge

Homi Bhabha was born on 30th October 1909 and had his school education at the Cathedral and John Connon High School in Bombay and his college education at the Elphinstone College and the Royal Institute of Science. At the age of 18, he left for England to pursue further studies at Cambridge. As desired by his parents, he completed his Mechanical Tripos with distinction and persuaded them to let him do a Mathematical Tripos since his own interests were in physics. Immediately after his second Tripos, he got a travelling fellowship and had the wonderful opportunity of working for short periods with Wolfgang Pauli at Zurich and with Enrico Fermi at Rome. In 1934, he was awarded the Isaac Newton Studentship at Cambridge which enabled him to complete his PhD under R H Fowler. He continued his research at Cambridge till 1939. The research that he did during this period had a direct bearing on the resolution of several important issues of cosmic ray phenomena and the interactions of particles especially electrons, protons and photons at high energies, in the context of the developments in the field of quantum mechanics and relativity. To appreciate these contributions of Bhabha it is necessary to become familiar with the status of cosmic ray studies in the early 30's.

Cosmic Rays in the Atmosphere: The 'Soft' and 'Penetrating' Components

The presence of a penetrating ionising radiation of extraterrestrial origin was established in 1912 by Victor Hess

Figure 1. Victor Hess in the Gondola in which he went up to an altitude of 16,000ft for measurement of cosmic ray intensity in 1912.

through a series of manned balloon experiments (*Figure 1*). The name 'Cosmic Rays' was given to this radiation by Millikan in 1925.

Analysis of the radiation at sea level and mountain altitude by a series of experiments with Geiger–Muller telescopes and magnetic cloud chambers, revealed that the radiation contained two components with distinctly different properties. One component, called the 'soft component', was easily absorbed in a few centimetres of lead, quite frequently multipled in number in passing through thin sheets of lead and also arrived at the observational level in multiples – as a shower of particles separated by several tens of centimetres. The second component, called 'penetrating component' could penetrate large thicknesses of matter, even a metre of lead, without multiplying. The only fundamental particles that were known at that time were the electrons, the photons, the protons and the α- particles. To this small list two more were added in 1932, the neutron and the positron. The positron was discovered by Anderson in a cloud chamber that had been set up to analyse the cosmic ray beam and its discovery was a great triumph to the relativistic quantum mechanical formulation of the theory of the electron developed by Dirac at Cambridge. Around the same time in the early 30's, Blackett and Occhialini who were also working at Cambridge had recorded several instances of multiple charged particles which had obvious non-ionising links between pairs of them. These events fitted beautifully the phenomenon of pair production or conversion of quanta into electron-positron pairs according to Dirac's theory. The calculations on the energy losses of charged particles by Bethe and Heitler revealed several surprising features – higher losses for lighter particles, (more for electrons than for protons of the same energy), higher losses in passing through matter of higher atomic number, and higher losses at higher energies.

All these features came in very handy in the explanation of cosmic ray anomalies. Clearly, Bhabha was at the right place at the right time. The very first paper of Bhabha entitled 'Zur Absorption der Hohenstrahlung' published in 1933 in *Zeitschrift Fur physik* was concerned with the explanation of the absorption features and shower production in cosmic rays. In 1936, Bhabha in collaboration with Heitler formulated the 'cascade theory of the electron' according to which a high energy electron passing through matter gave rise to a high energy photon by bremstrahlung process and the photon in turn produced a pair of positive and negative electrons; these in turn led to further production of photons and the cascade process continued until the energy of the particles fell below a critical value. Carlson and Oppenheimer also developed a similar theory simultaneously in the USA. Based on Bethe–Heitler cross sections, Bhabha and Heitler made quantitative estimates of the number of electrons in the cascade at different depths, for different initiating energy of the electrons. These calculations agreed with the experimental findings of Bruno Rossi in cosmic ray showers. The problem of the 'soft' component was thus totally resolved.

Bhabha and Heitler made quantitative estimates of the number of electrons in the cascade at different depths, for different initiating energy of the electrons.

In a classic paper entitled 'On the penetrating component of cosmic radiation' communicated to the *Proceedings of the Royal Society* in July 1937, Bhabha made a careful analysis of the experimental data on the soft and penetrating components and concluded that a 'breakdown' of the quantum mechanical theory of radiation at higher energies as proposed by some theorists would not explain the experimental results on the latitude effect of cosmic rays and the shape of the transition curve of large cosmic ray bursts. He emphasised that these features would find a natural explanation if cosmic radiation contained charged particles of mass intermediate between electron and proton and set the mass as ~100 electron masses.

Around the same time, Neddermeyer and Anderson, and Street and Stevenson discovered in their cloud chamber experiments, charged particles of intermediate mass whose mass was set at ~ 200 electron masses. The name 'meson' was given to this new particle.

Bhabha predicted (in a paper in *Nature*, 1938) that the meson would be unstable and would probably decay into an electron and neutrino. The phenomenon of meson decay was helpful in resolving anomalous absorption of the penetrating component in the atmosphere. The relativistic elongation of time as predicted by the special theory of relativity was confirmed through meson decay experiments.

Bhabha and Cosmic Ray Research at the Indian Institute of Science, Bangalore

Bhabha came on a brief holiday to India in 1939. He could not go back to England as planned, since the second world war broke out in September 1939, and there was the prospect of heavy bombing over England by the Germans. Bhabha decided to stay back in India for a while. This decision turned out to be a turning point, a landmark not only in the academic career of Bhabha, but also in the advancement of Indian science and technology in the post independent era.

Bhabha joined the Physics Department of the Indian Institute of Science, headed by C V Raman. He got a special grant from the Sir Dorab Tata Trust. He gathered some students to work with him in theoretical particle physics and one of them was Harish-Chandra, who later held a professorial chair in mathematics at the Princeton Institute of Advanced Studies.

Bhabha predicted (in a paper in *Nature*, 1938) that the meson would be unstable and would probably decay into an electron and neutrino.

In parallel, Bhabha also started experimental work in cosmic rays. He was cognisant of the unique advantages of India to work in this field – wide range of latitudes from equator in the south to 25° N in Kashmir within the boundaries of a single country; mountain stations in the south and north and the deepest mines in the world. Millikan had come all the way from the USA to do experiments at several stations in India in the mid 30's.

With a uniquely designed GM telescope, which Bhabha built with the help of S V C Iya, the penetrating particle intensities were measured at altitudes of 5000, 10,000, 15,000, 20,000, 25,000 and 30,000 ft, using a B-29 bomber aircraft belonging to the US Air Force. These constituted the first measurements at such high altitudes in an equatorial latitude. Comparison with the measurements of Schein, Jesse and Wollan in the USA, established that no marked increase of intensity occurred between 3.3°N and 52°N even at an altitude of 30,000ft., in contrast to the total intensity which exhibited very pronounced latitude effect at such altitudes.

At the Indian Institute of Science, Bhabha also got constructed a 12" diameter cloud chamber identical to the one operating in Manchester. R L Sengupta, who had worked in Blackett's Laboratory helped Bhabha in the design and construction of this chamber, which was used by M S Sinha to study the scattering characteristics of mesons. Vikram Sarabhai set up a telescope to study the time variation of cosmic ray intensity.

Bhabha and Cosmic Ray Research at the Tata Institute of Fundamental Research

While at the Indian Institute of Science, Bhabha recognised the need for setting up in the country an institute solely devoted to the pursuit of fundamental research especially in the area of nuclear science that was emerging as a virgin area of fundamental science. The developments in the field of cosmic ray studies and in the area of nuclear physics with accelerators had convinced Bhabha that the future lay in these areas. With financial support from the Sir Dorab Tata Trust and the Government of Maharashtra, Bhabha established the Tata Institute of Fundamental Research in Bombay in June 1945. The formal inauguration of TIFR was at IISc, Bangalore. The TIFR became an aided institution under the Department

With financial support from the Sir Dorabji Tata Trust and the Government of Maharashtra, Bhabha established the Tata Institute of Fundamental Research in Bombay in June 1945.

of Atomic Energy later and was recognised as the National Centre for Nuclear Science and Mathematics by the Government of India. Bhabha himself used to say that TIFR was the 'cradle of the Atomic Energy Programme' of the country. Bhabha was the Director of TIFR from 1945 to January 1966 – till his untimely death in a tragic air crash on the Alps.

The TIFR naturally started with a major experimental programme in cosmic rays, taking cognisance of the fact that cosmic ray research had entered its second phase the world over. The π-meson as the parent of the μ-meson was discovered in 1947 by Powell and his collaborators at the University of Bristol exposing the newly developed high sensitivity nuclear emulsions in the Jungfraujoch mountains in Switzerland. Rochester and Butler discovered the same year the V^0 particles, which were later identified as the K-mesons and hyperons through nuclear emulsion experiments by several groups. Bradt and Peters discovered around the same time the presence of α-particles and other stripped heavy nuclei in the primary cosmic radiation which consisted predominantly of protons. Also, most importantly, the act of meson production had been caught both in nuclear emulsions and in multiplate cloud chambers.

With these developments, the new directions of cosmic ray research had become clear. To enter the international arena in this field, the emphasis had to be on: (i) the investigations on the primary component – spectrum, composition, anisotropy of arrival directions; relative proportions of rare nuclei, electrons, γ-rays; (ii) the detailed study of the characteristics of nuclear collisions of the primaries as well as of the secondaries produced in these collisions; (iii) the studies on the penetrating components – muons and neutrinos in deep underground installations; (iv) studies on the extensive air showers initiated in the atmosphere through the nuclear and electromagnetic cascades by the primaries; (v) study of the radio isotopes produced by cosmic rays; (vi) time variation studies on cosmic ray intensity and correlations with solar activity (*Figure 2*).

These multidimensional studies to be carried out in a variety of locations with specially designed detector systems, required the inhouse development of a variety of technologies – to name a few – plastic balloon fabrication technology, fabrication of GM counters, plastic scinti-llators, multiplate cloud chambers, pulsed electronic circuits and even a digital computer. Thanks to the organisational genius of Bhabha, all this was done in a record time in TIFR itself. The Indian industry was very backward in the 40's and 50's and importing then was just not thinkable because of shortage of foreign exchange and the enormous delays of transportation. The cosmic ray programme did get a fillip in the 50's by Bernard Peters, the co-discoverer of heavy primaries and M G K Menon who worked for 8 years in Powell's Laboratory, joining the Nuclear Emulsion Group of TIFR.

Figure 2. Bhabha and A S Rao with a typical cosmic ray telescope – of the type that was being launched from the Central College grounds, Bangalore in the late forties, on clusters of rubber balloons.

At the International Conference on Cosmic Rays held at Bagneres in 1953, TIFR made its first impact by presenting very significant results on K-mesons and hyperons obtained from the analysis of emulsion stacks exposed at Hyderabad. The emulsion group kept a high profile of original contributions in the field of high energy interaction studies, the relative abundances of Li, Be and B in the primaries, hyperfragments and on the spectrum of primary electrons. The deep underground experiments in the Kolar Gold Fields initiated at the instance of Bhabha as early as 1950, and which continued for more than four decades, till 1994, was another line of activity in which pioneering contributions were made – most accurate μ-meson intensity and angular distribution measurements upto very high energies (*Figure 3*), detection of neutrino induced interactions with a visual detector, limits on the lifetime of protons etc. These involved very large scale installations and also international collaborations. Extensive air shower array with a variety of detectors for different components – scintillators, Cerenkov counters, total absorption spectrometer, multiplate cloud chamber started operating in the late 50's in

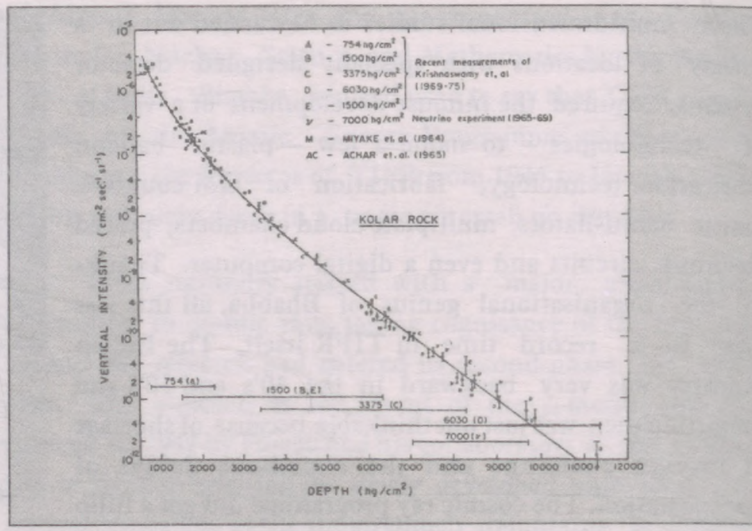


Figure 3. The variation of intensity of penetrating particles as a function of depth-based on a variety of experiments at the Kolar Gold Fields.

the mountain station at Ooty. The time structure measurements of hadrons with the total absorption spectrometer, led to the first recognition of increased cross section for the production of nucleons and antinucleons at high energies. Bhabha, when he visited the Ooty Laboratory in 1964, was thrilled to see the world's largest multiplate cloud chamber operating there. This cloud chamber gave unique information on the highest energy jets produced by the incidence of several parallel hadrons (*Figure 4*). At the Kolar Gold Fields, a second air shower array was set up at the surface of the mines with large area detectors at several depths underground that recorded the associated very high energy muons. This set-up gave very valuable information on the composition of the primaries in the crucial

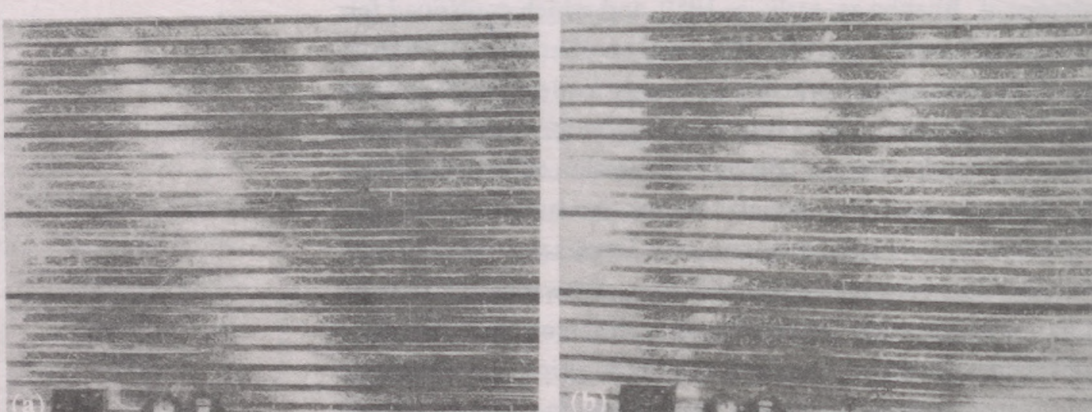


Figure 4. Development of cascade in a multiplate cloud chamber at Ooty. (a) A cascade having an elongated tube-like structure which is not completely absorbed even after 20 radiation lengths, the estimated energy is 2.4TeV. (b) A cascade which develops from the first plate of the chamber and shows a rapid absorption after the maxima. The method of cascade widths has been used for energy estimation which is 750GeV.

knee region 10^{14}–10^{16}ev.

In a short article like this it is difficult to do full justice and bring out the full flavour and ramification of all the work in cosmic rays that got initiated at the instance of Bhabha. Bhabha's was a multidimensional, many splendoured personality that influenced not only cosmic ray research, but many other fields too. But cosmic rays were very dear to him, all through his life, may be because his very first paper was on cosmic rays.

Even 86 years after the discovery of cosmic rays, 50 years after entering the second phase, despite colossal efforts by groups all over the world, not a single source of cosmic rays of high energy (> 20 Gev) has been identified even though it is firmly established that the spectrum extends beyond 10^{20}ev. The mechanism by which particles are accelerated to such high energies is also not known. The high rotating magnetic field environments of the neutron stars in pulsars in the galaxy and the extragalactic AGN (active galactic nuclei) with suspected giant blackholes in their centres are thought to be the strongest candidates. Gigantic multiplex installations are coming up to settle this question. What other exotic particles there are among the primaries and what new particles are produced in super high energy collisions are other aspects which are receiving special attention in the design of next generation cosmic ray experiments.

Suggested Reading

[1] Bruno Rossi. *Cosmic Rays*. McGraw-Hill Book Company, 1964.

[2] B V Sreekantan. Cosmic Ray Physics and Astrophysics. *Proc. Indian Natl. Sci. Acad.* Vol.44, 1975.

[3] *Homi Jehangir Bhabha – Collected Papers*. Eds. B V Sreekantan, Virendra Singh, B M Udgaonkar. Tata Institute of Fundamental Research, 1985.

Address for Correspondence
B V Sreekantan
National Institute of Advanced Studies
Indian Institute of Science Campus, Bangalore 560 012, India.
Email: bvsreek@tifrbng.ncbs.res.in

Homi J Bhabha in his address to the International Council of Scientific Unions on January 7, 1966.

An important question which we must consider is whether it is possible to transform the economy of a country to one based on modern technology developed elsewhere without at the same time establishing modern science in the country as a live and vital force ... the problem of establishing science as a live and vital force in society is an inseparable part of the problem of transforming an industrially underdeveloped to a developed country.

The Class Number Problem

2. An Introduction to Algebraic Number Theory

Rajat Tandon

Rajat Tandon received his PhD from Yale University, USA. After an initial period as a visiting post-doctoral fellow at TIFR, he joined the faculty of the North Eastern Hill University, where he taught for a few years. For about two decades now, he has been with the University of Hyderabad. His interests are in the area of number theory.

The second part gives an introduction to 'algebraic number theory', defines class numbers for finite extensions of the field of rational numbers and proves that in the context of quadratic fields, this definition coincides with the definition of class numbers via binary quadratic forms given in the first part [1].

We have seen in the first part of this article in the previous issue that some seemingly innocuous questions starting with the formula $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$ lead to fairly deep mathematics. This is typical of the subject. It is so important to ask the right question – " ask an impertinent question and you get a pertinent answer ".

The roots of $aX^2+bX+c=0$ are given by $\frac{-b\pm\sqrt{\Delta}}{2a}$ where $\Delta = (b^2-4ac)$i.e., they are of the form $x+y\sqrt{\Delta}$ with x and y rational. The set $\mathbf{Q}(\sqrt{\Delta})$ of elements of the form $x+y\sqrt{\Delta}$ with x and y rational forms a subfield of the field of complex numbers, $\mathbf{C}$. $\mathbf{Q}(\sqrt{\Delta})$ is also a vector space over the rationals if we define scalar multiplication by $\lambda(x+y\sqrt{\Delta}) = \lambda x+\lambda y\sqrt{\Delta}$. $\{1, \sqrt{\Delta}\}$ is a basis of $\mathbf{Q}(\sqrt{\Delta})$ over $\mathbf{Q}$, and $\mathbf{Q}$ is a subfield of $\mathbf{Q}(\sqrt{\Delta})$. This process can easily be generalised. For instance, let p be a prime and $\zeta = e^{2\pi i/p}$. Let $\mathbf{Q}(\zeta)$ be the set of complex numbers of the form $x_0+x_1\zeta+x_2\zeta^2+\cdots+x_{p-2}\zeta^{p-2}$ with x_i rational. Note that $1+\zeta+\zeta^2+\zeta^3+\cdots+\zeta^{p-1} = 0$ so ζ^{p-1} can be written in terms of $1, \zeta, \zeta^2, \zeta^3, \ldots, \zeta^{p-2}$. Check that $\mathbf{Q}(\zeta)$ is a subfield of $\mathbf{C}$ containing $\mathbf{Q}$ and that $1, \zeta, \zeta^2, \ldots, \zeta^{p-2}$ is a basis of $\mathbf{Q}(\zeta)$ over $\mathbf{Q}$ with scalar multiplication being defined in the obvious way. These are examples of fields containing $\mathbf{Q}$ which are finite dimensional as vector spaces over $\mathbf{Q}$. Such fields are known as *algebraic number fields* and were the object of detailed study by Dedekind, Kronecker and Kummer in the 19th century. Amongst the

[1] See *Resonance*, Vol.3, No.6, 1998.

several motivations for studying such fields were three problems suggested by Greek geometers:
(i) To trisect any given angle
(ii) To construct a cube whose volume is twice that of a given cube.
(iii) To construct a square equal in area to a given circle.

These constructions were to be done by 'ruler and compass only' in the manner that we are taught at school. The second problem boils down to being able to construct by ruler and compass the real root of $X^3 - 2$. Galois and Abel looked at such problems and their work gave a huge impetus to the systematisation of algebra and algebraic number theory.

The examples given above, $\mathbf{Q}(\sqrt{\Delta})$ and $\mathbf{Q}(\zeta)$, have been generated by single elements ($\sqrt{\Delta}$ and ζ) which satisfy some polynomial with rational (in fact, integral) coefficients ($X^2 - \Delta, X^p - 1$ respectively). Indeed, it can be shown that any subfield of $\mathbf{C}$ containing $\mathbf{Q}$ which is n-dimensional as a vector space over $\mathbf{Q}$ consists of elements of the form $x_0 + x_1\alpha + x_2\alpha^2 + \cdots + x_{n-1}\alpha^{n-1}$ where the x_i are rationals and α is a complex number which satisfies a polynomial equation of degree n with rational coefficients.

The first thing we would want to know about such fields is whether they have a subring in them in much the same way that $\mathbf{Q}$ contains $\mathbf{Z}$ and every element of $\mathbf{Q}$ is a ratio of two (one non-zero) elements of $\mathbf{Z}$. One 'natural' possibility in $\mathbf{Q}(\sqrt{\Delta})$ could be $\mathbf{Z} + \mathbf{Z}\sqrt{\Delta}$, i.e., elements of the form $a + b\sqrt{\Delta}$ with a and b integers or in other words $\mathbf{Z}$-linear combinations of the basis 1, $\sqrt{\Delta}$. Similarly one could consider $\mathbf{Z}+\mathbf{Z}\zeta+\mathbf{Z}\zeta^2+\mathbf{Z}\zeta^3+\cdots \mathbf{Z}\zeta^{p-2}$ in $\mathbf{Q}(\zeta)$. But immediately one would recognise a difficulty in basing a definition which depends on the choice of a basis. For instance, $\mathbf{Q}(\sqrt{\Delta}) = \mathbf{Q}(\sqrt{4\Delta})$ but $\mathbf{Z} + \mathbf{Z}\sqrt{\Delta} \neq \mathbf{Z} + \mathbf{Z}\sqrt{4\Delta}$ or observe that if $p = 3$ then $\zeta = e^{2\pi i/3} = \frac{-1+\sqrt{-3}}{2}$ so $\mathbf{Q}(\zeta) = \mathbf{Q}(\sqrt{-3})$ but $\mathbf{Z} + \mathbf{Z}\zeta \neq \mathbf{Z} + \mathbf{Z}\sqrt{-3}$. To get around the problem of square factors of Δ we will henceforth assume that Δ is a fundamental discriminant see Part I of this article . Hence the only square factor Δ can have is 4.

We have already seen that the fields above are generated by elements which satisfy a monic (leading coefficient 1) polynomial with rational coefficients. In fact, every element $a+b\sqrt{\Delta}$ in $\mathbf{Q}(\sqrt{\Delta})$ satisfies the polynomial $X^2-2aX+(a^2-b^2\Delta)=0$. This suggests an alternative. Why not consider those elements of $\mathbf{Q}(\sqrt{\Delta})$ (or $\mathbf{Q}(\zeta)$) which satisfy a monic polynomial with coefficients in $\mathbf{Z}$? Such elements are called *algebraic integers* (in the given field). Do such elements form a subring I, i.e., are they closed under addition and multiplication? The answer is 'yes'. Observe that $a+b\sqrt{\Delta}$ will be an element of the given type provided $2a\in\mathbf{Z}$ and $a^2-b^2\Delta\in\mathbf{Z}$. Suppose then that $a+b\sqrt{\Delta}$ and $c+d\sqrt{\Delta}$ are such that $2a,2c\in\mathbf{Z}$ and $a^2-b^2\Delta, c^2-d^2\Delta\in\mathbf{Z}$. Observe that $2(a+c)\in\mathbf{Z}$ and $(a+c)^2-(b+d)^2\Delta=(a^2-b^2\Delta)+(c^2-d^2\Delta)+2ac-2bd\Delta$. We say that a rational number is a half integer if it is of the form $\frac{l}{2}$ where l is odd. We make the following observations which can easily be proved by the reader: for $a,b\in\mathbf{Q}$, $2a$ and $a^2-b^2\Delta$ are integers implies

(i) $2b\in\mathbf{Z}$ since Δ has no square free factor other than possibly 4;

(ii) if Δ is even then a must be an integer and b either an integer or half integer;

(iii) if Δ is odd a and b must be either both integers or both half integers.

In all cases it can then be seen that if $2a,2c\in\mathbf{Z}$ and $a^2-b^2\Delta, c^2-d^2\Delta\in\mathbf{Z}$ then $2ac-2bd\Delta\in\mathbf{Z}$ and therefore that $(a+c)^2-(b+d)^2\Delta\in\mathbf{Z}$. On the other hand

$$\begin{aligned}(a+b\sqrt{\Delta})\cdot(c+d\sqrt{\Delta}) &= ac+bd\Delta+(ad+bc)\sqrt{\Delta}\\ (ac+bd\Delta)^2-(ad+bc)^2\Delta &= (a^2-b^2\Delta)\cdot(c^2-d^2\Delta)\end{aligned}$$

are both in $\mathbf{Z}$. Hence I is indeed closed under addition and multiplication.

Exercise: Show that

(i) in $\mathbf{Q}(\sqrt{-1})$ we have $I=\{a+b\sqrt{-1}\,|\,a,b\in\mathbf{Z}\}$

(ii) in $\mathbf{Q}(\sqrt{-3})$ $I=\{\frac{a+b\sqrt{-3}}{2}\,|\,a,b\in\mathbf{Z}\; a\equiv b \pmod 2\}=\mathbf{Z}+\mathbf{Z}\zeta$ where $\zeta=\frac{-1+\sqrt{-3}}{2}$ is a cube root of unity.

Would every element of $\mathbf{Q}(\sqrt{\Delta})$ be a ratio of two elements of

I? We note that $\mathbf{Z}+\mathbf{Z}\sqrt{\Delta} \subseteq I$ and $\frac{a}{b}+\frac{c}{d}\sqrt{\Delta} = \frac{ad+bc\sqrt{\Delta}}{bd}$ so this is trivially true. What other properties of $\mathbf{Z}$ would we like I to have? The best would be unique factorisation. In $\mathbf{Z}$ we have the notion of a prime number and we know that every number can be written upto sign uniquely as a product of distinct prime powers viz $n = \pm p_1^{e_1} p_2^{e_2} \ldots p_r^{e_r}$ where the p_i are distinct primes and, moreover, if n is also equal to $\pm q_1^{f_1} q_2^{f_2} \ldots q_s^{f_s}$ then after changing the order of the q_i's, if necessary, we have $r = s, p_i = q_i$ and $e_i = f_i$ for all i.

Imagine the usefulness of having such a property in I. For instance consider $\mathbf{Q}(\zeta)$ as above and the ring of integers I in $\mathbf{Q}(\zeta)$,i.e., the set of all elements in $\mathbf{Q}(\zeta)$ which satisfy a monic polynomial in $\mathbf{Z}[X]$, the ring of polynomials in one variable with integer coefficients. Suppose there exist non-zero integers x, y, z such that $x^p + y^p = z^p$. Then

$$x^p = z^p - y^p = (z-y)(z-\zeta y)(z-\zeta^2 y)\ldots(z-\zeta^{p-1}y). \quad (*)$$

It is easy to see that $x \in I$ and $z - \zeta^i y \in I$. If we have unique factorisation in I there is just a chance that $(*)$ may give us a contradiction to unique factorisation (or allow us to use the method of descent) and we may prove Fermat's [1] last theorem! It is just possible that Fermat had some such proof in mind when he wrote in the margin ....

We would first need the notion of a prime element in I. This is accomplished more or less as in $\mathbf{Z}$ – negatives allowed. So we consider $-2, -3, -5, \ldots$ also as primes.

Definition 1: An integer n is a prime if whenever n is written as a product ab of two integers then either a or b must be ± 1. Note that ± 1 are the only units in $\mathbf{Z}$, i.e. elements in $\mathbf{Z}$ with a multiplicative inverse.

There is another way of defining a prime number.

Definition 1′: An integer $p \neq \pm 1$ is a prime if and only if whenever p divides a product of integers ab then p must divide either a or b.

Recall if n is an integer then $n\mathbf{Z}$, the set of multiples of n, forms an ideal in $\mathbf{Z}$ (an ideal J in a commutative ring R is

[1] Incidentally, this is what Gauss had to say about FLT. "I confess that Fermat's theorem as an isolated proposition has very little interest for me because I could easily lay down a multitude of such propositions which one could neither prove nor dispose off." Gauss said that FLT had induced him to recall some of his earlier ideas in higher arithmetic but that he was not in a position to go back to that work because of his circumstances. "Still I am convinced that if I am as lucky as I dare hope and if I succeed in taking some of the principal steps in that theory, then Fermat's theorem will appear as only one of the least interesting corollaries."

an additive subgroup of R which has the property: $x \in J$, $r \in R$ implies $rx \in J$). If an ideal I in a ring satisfies the property: $ab \in I$ implies either $a \in I$ or $b \in I$ it is called a prime ideal. So the integer p is a prime number is the same thing as saying that $p\mathbf{Z}$ is a prime ideal in $\mathbf{Z}$. It is easy to see that the two definitions we have given are equivalent in $\mathbf{Z}$.

Based on the above we could define in an arbitrary commutative ring with unity R (all our rings will be so) an element π to be prime either by requiring that whenever $\pi = ab$ either a or b must be a unit in R or by requiring that the ideal πR, consisting of all multiples of π, is a prime ideal. Unfortunately in an arbitrary ring the two definitions are not equivalent. An element π which satisfies the first property is said to be *irreducible* whereas if πR is a prime ideal we call π a *prime.* In integral domains (commutative rings with no zero divisors) all primes are irreducible but not vice-versa.(Exercise: Prove this.)

A domain in which every non-zero non-unit can be written as a product of irreducibles in an essentially unique way, that is upto order and multiplication by units ($6 = 2 \cdot 3 = 3 \cdot 2 = (-2) \cdot (-3) = (-3) \cdot (-2)$) is called a *unique factorisation domain* (UFD). Clearly, $\mathbf{Z}$ is a UFD and it is easy to check that $J = \mathbf{Z} + \mathbf{Z}i$ is also a UFD.

$\mathbf{Z}$ has another property which is somewhat stronger – every ideal in $\mathbf{Z}$ is of the form $n\mathbf{Z}$ where n is an integer. A domain D which has the property that every ideal in it is of the form xD for some x in D is called a *Principal Ideal Domain* (PID) and every PID is a UFD. If we could show that the ring of integers I in an algebraic number field is always a PID then we could use the argument given above for FLT. Unfortunately I is not always a PID. For instance, consider $\mathbf{Q}(\sqrt{-20})$; then $I = \mathbf{Z} + \mathbf{Z}\sqrt{-5}$ and we have $6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$. It is easy to check that 2, 3, $1 \pm \sqrt{-5}$ are all irreducible elements in I. We remark that the ring of integers of an algebraic number field is a UFD if and only if it is a PID.

Recall that if $\mathcal{A}$ and $\mathcal{B}$ are two ideals in a ring R then we define their product $\mathcal{A} \cdot \mathcal{B} = \{\sum_{i=1}^{i=n} a_i b_i | a_i \in \mathcal{A}, b_i \in \mathcal{B}$, for some $n\}$. This is also an ideal. Though I is not always a PID it is true that every ideal in I can be written uniquely, except for order, as a product of prime ideals. This gives us the first hint that the concept of an ideal may be at least as important as the notion of an element. Note that in a PID the two notions are almost the same as every ideal is generated by a single element which is uniquely determined upto units.

So if I is not always a PID then how 'bad' is it? The set $\mathcal{I}$ of ideals in I under the product defined above form a semigroup (I itself is the identity). We define an equivalence relation on this set $\mathcal{I}$ as follows: $\mathcal{A} \sim \mathcal{B}$ if there exist $\alpha, \beta \in I$ such that $\alpha I \cdot \mathcal{A} = \beta \mathcal{I} \cdot \mathcal{B}$. It is easy to check that this gives us an equivalence relation on $\mathcal{I}$ and the product on $\mathcal{I}$ induces a product on the set of equivalence classes $\mathcal{I}/\sim$: $[\mathcal{A}] \cdot [\mathcal{B}] = [\mathcal{A} \cdot \mathcal{B}]$. The crucial point here is to check that '$\cdot$' as defined above is well defined, i.e., if $\mathcal{A} \sim \mathcal{A}'$ and $\mathcal{B} \sim \mathcal{B}'$ then $\mathcal{A} \cdot \mathcal{B} \sim \mathcal{A}' \cdot \mathcal{B}'$. The set of equivalence classes $\mathcal{I}/\sim$ with this product is actually a group. It is one of the fundamental theorems of algebraic number theory that this group is finite – not just for quadratic extensions of $\mathbf{Q}$ but for any finite extension of $\mathbf{Q}$. The order of this group is called the *class number* of the extension. The class number of $\mathbf{Q}(\sqrt{\Delta})$ will be denoted by $h'(\Delta)$. Note that the class number is one if and only if I is a PID.

Let now Δ be a negative fundamental discriminant,i.e., a negative integer Δ which is congruent to 0 or 1 modulus 4 and which cannot be written in the form $\Delta_0 n^2$ where Δ_0 is another discriminant and n is an integer. Hence 4 is the only possible square factor of Δ. Recall that we have defined $h(\Delta)$ to be the number of equivalence classes of primitive binary integral quadratic forms. Remarkably:

THEOREM: $h(\Delta) = h'(\Delta)$.

In order to prove this we first observe that if $\alpha = a + b\sqrt{\Delta}$ is in the ring of integers I of $\mathbf{Q}(\sqrt{\Delta})$ then so also is $\overline{\alpha} =$

$a - b\sqrt{\Delta}$. Hence so also is $\alpha\overline{\alpha}$ which is an integer. Hence if $\mathcal{A}$ is any non-zero ideal of I then $\mathcal{A} \cap \mathbf{Z} \neq (0)$. Clearly $\mathcal{A} \cap \mathbf{Z}$ is an ideal in $\mathbf{Z}$ so $\mathcal{A} \cap \mathbf{Z} = a\mathbf{Z}$ for some integer $a > 0$. Observe also that any non-zero ideal of I cannot be contained in $\mathbf{Z}$.

In order to make life a bit easier we will assume in what follows that Δ is odd and hence that $I = \mathbf{Z} + \mathbf{Z} \cdot \frac{1+\sqrt{\Delta}}{2}$ (proof?). Let $\mathcal{A}$ be an ideal in I. Define

$$J = \{r \in \mathbf{Z} | r \cdot \frac{1+\sqrt{\Delta}}{2} + s \in \mathcal{A}\}$$

for some $s \in \mathbf{Z}$. Then J is an ideal in $\mathbf{Z}$ and since $\mathcal{A} \not\subseteq \mathbf{Z}, J$ is non-zero. Let $J = t\mathbf{Z}$, $t > 0$. Then there exists an $s \in \mathbf{Z}$ such that $t\frac{1+\sqrt{\Delta}}{2} + s \in \mathcal{A}$. We claim that $\mathcal{A} = a\mathbf{Z} + \frac{(t+2s)+t\sqrt{\Delta}}{2}\mathbf{Z}$. Clearly the right hand side is contained in $\mathcal{A}$. Let $\alpha = u + v\frac{1+\sqrt{\Delta}}{2} \in \mathcal{A}$. Then $v \in J$ so $v = tv'$ for some $v' \in \mathbf{Z}$. Therefore

$$\begin{aligned} \alpha - v'\frac{(t+2s)+t\sqrt{\Delta}}{2} &= u + tv'\frac{1+\sqrt{\Delta}}{2} - v'\frac{(t+2s)+t\sqrt{\Delta}}{2} \\ &= u - sv' \in \mathcal{A} \cap \mathbf{Z} = a\mathbf{Z}. \end{aligned}$$

Therefore $\alpha \in a\mathbf{Z} + \frac{(t+2s)+t\sqrt{\Delta}}{2}\mathbf{Z}$. Hence every ideal $\mathcal{A}$ in I is of the form $a\mathbf{Z} + \frac{b+c\sqrt{\Delta}}{2}\mathbf{Z}$, $a > 0, c > 0$. For this to be an ideal it must be closed under multiplication by $\frac{1+\sqrt{\Delta}}{2}$. Hence $a \cdot \frac{1+\sqrt{\Delta}}{2} \in a\mathbf{Z} + \frac{b+c\sqrt{\Delta}}{2}\mathbf{Z}$,i.e., there exist integers m, n such that $a \cdot \frac{1+\sqrt{\Delta}}{2} = ma + n \cdot \frac{b+c\sqrt{\Delta}}{2} \Longrightarrow a = nc$ and $1 = 2m + \frac{b}{c}$ i.e., c divides a, c divides b and $\frac{b}{c}$ is odd. Let $a = tc, b = uc, u$ odd. Then $a\mathbf{Z} + \frac{b+c\sqrt{\Delta}}{2}\mathbf{Z} = tc\mathbf{Z} + \frac{uc+c\sqrt{\Delta}}{2}\mathbf{Z} = c[t\mathbf{Z} + \frac{u+\sqrt{\Delta}}{2}\mathbf{Z}]$. Hence, every ideal $\mathcal{A}$ in I is of the form $c[t\mathbf{Z} + \frac{u+\sqrt{\Delta}}{2}\mathbf{Z}]$, with $c > 0, t > 0$ and u odd. Further, again since $\mathcal{A}$ is closed under multiplication by $\frac{1+\sqrt{\Delta}}{2}$, $c \cdot \frac{u+\sqrt{\Delta}}{2} \cdot \frac{1+\sqrt{\Delta}}{2} \in \mathcal{A}$. Hence there exist integers h, k such that $\frac{(u+\Delta)+(1+u)\sqrt{\Delta}}{4} = ht + k.\frac{u+\sqrt{\Delta}}{2}$. Therefore, $k = \frac{1+u}{2}$ and $\frac{u+\Delta}{4} = ht + \frac{ku}{2} = ht + \frac{u(1+u)}{4}$. Hence $\Delta = u^2 + 4ht$. We have proved:

Proposition: Every ideal in I is of the form $t(a\mathbf{Z}+\frac{b+\sqrt{\Delta}}{2}\mathbf{Z})$ for some integers a, b, t with $t > 0, a > 0$ and such that there exists an integer c with $\Delta = b^2 - 4ac$.

Proof of the THEOREM: We denote by $[aX^2 + bXY + cY^2]$ the equivalence class of the form $aX^2 + bXY + cY^2$ in $S_1(\Delta)$. We denote by $[\mathcal{A}]$ the equivalence class of the ideal $\mathcal{A}$ in $\mathcal{I}$. Define

$$\epsilon : S_1(\Delta)/\sim \;\longrightarrow\; \mathcal{I}/\sim$$
$$[aX^2 + bXY + cY^2] \;\longmapsto\; [a\mathbf{Z} + \frac{b+\sqrt{\Delta}}{2}\mathbf{Z}].$$

Then the proposition we have proved above shows that ϵ is surjective. We need, of course, to show that ϵ is well defined. For this we must show that if

$$A \cdot (aX^2 + bXY + cY^2) = a'X^2 + b'XY + c'Y^2$$

where A is either $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ then $a\mathbf{Z}+\frac{b+\sqrt{\Delta}}{2}\mathbf{Z} \sim a'\mathbf{Z} + \frac{b'+\sqrt{\Delta}}{2}\mathbf{Z}$.

If $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ then $a' = a$ and $b' = b + 2a$ which implies that $a\mathbf{Z} + \frac{b+\sqrt{\Delta}}{2}\mathbf{Z} = a'\mathbf{Z} + \frac{b'+\sqrt{\Delta}}{2}\mathbf{Z}$.

If $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ then $a' = c$ and $b' = -b$ so

$$a'\mathbf{Z}+\frac{b'+\sqrt{\Delta}}{2}\mathbf{Z} = c\mathbf{Z}+\frac{-b+\sqrt{\Delta}}{2}\mathbf{Z} = \frac{b^2-\Delta}{4a}\mathbf{Z}+\frac{-b+\sqrt{\Delta}}{2}\mathbf{Z}.$$

Therefore $a(a'\mathbf{Z} + \frac{b'+\sqrt{\Delta}}{2}\mathbf{Z}) = \frac{(-b+\sqrt{\Delta})}{2}.(a\mathbf{Z} + \frac{b+\sqrt{\Delta}}{2}\mathbf{Z})$ and we have proved what was required.

In order to prove our theorem we must show that ϵ is a bijection. Only the injectivity of ϵ is left. Before proving injectivity we make two remarks:

(a) If $\mathcal{A}$ and $\mathcal{B}$ are two ideals in I then they are equivalent if there exists $\alpha, \beta \in I$ such that $\alpha.\mathcal{A} = \beta.\mathcal{B}$. But this is equivalent to $\alpha\overline{\alpha}\mathcal{A} = \overline{\alpha}\beta\mathcal{B}$ and $\alpha\overline{\alpha}$ is a positive integer.

Hence $\mathcal{A} \sim \mathcal{B}$ if and only if there exists an integer $t > 0$ and $\beta \in I$ such that $t \cdot \mathcal{A} = \beta \cdot \mathcal{B}$.
(b) If $\alpha, \beta \in I$ and $\alpha\mathbf{Z} + \beta\mathbf{Z} = \gamma\mathbf{Z} + \delta\mathbf{Z}$ then there exists an integral 2×2 matrix A of determinant ± 1 such that $A \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \gamma \\ \delta \end{pmatrix}$

Suppose now that

$$\epsilon([aX^2 + bXY + cY^2]) = \epsilon([a'X^2 + b'XY + c'Y^2])$$

i.e.,

$$a\mathbf{Z} + \frac{b + \sqrt{\Delta}}{2}\mathbf{Z} \sim a'\mathbf{Z} + \frac{b' + \sqrt{\Delta}}{2}\mathbf{Z}.$$

Hence there exists an integer $t' > 0$ and $\alpha = \frac{p+q\sqrt{\Delta}}{2}$ in I such that $\alpha \cdot (a\mathbf{Z} + \frac{b+\sqrt{\Delta}}{2}\mathbf{Z}) = t' \cdot (a'\mathbf{Z} + \frac{b'+\sqrt{\Delta}}{2}\mathbf{Z}) = \mathcal{A}$(say). We must show that

$$a'X^2 + b'XY + c'Y^2 = A \cdot (aX^2 + bXY + cY^2)$$

for some A in SL(2,$\mathbf{Z}$).

Case 1: Let $q = 0$ and $t = p/2$. Then $at\mathbf{Z} = a't'\mathbf{Z} = \mathcal{A} \cap \mathbf{Z}$. We may without loss of generality assume that $at = a't'$ and hence $t > 0$. There exist integers m, n such that $t \cdot \frac{b+\sqrt{\Delta}}{2} = ma't' + nt' \cdot \frac{b'+\sqrt{\Delta}}{2}$ which implies that $t = nt'$ and hence $a' = na$. There also exist integers k, l such that $t'\frac{b'+\sqrt{\Delta}}{2} = kta + lt\frac{b+\sqrt{\Delta}}{2}$. Hence $ln = 1$ or $n = 1, t = t', a = a'$ and $b' = b + 2ak$. It is now easy to see that

$$a'X^2 + b'XY + c'Y^2 = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \cdot (aX^2 + bXY + cY^2).$$

Case 2: $(q \neq 0)$ In view of case 1 we may assume that $(p, q) = 1$. By the proposition above and remark (b) there exists an integral matrix $A = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$ of determinant ± 1 such that

$$A \cdot \begin{pmatrix} a\frac{p+q\sqrt{\Delta}}{2} \\ (\frac{b+\sqrt{\Delta}}{2}) \cdot (\frac{p+q\sqrt{\Delta}}{2}) \end{pmatrix} = \begin{pmatrix} t'a' \\ t' \cdot \frac{b'+\sqrt{\Delta}}{2} \end{pmatrix}$$

or, in fact, by multiplying by the matrix $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, if necessary, we can assume that A is in SL(2,$\mathbf{Z}$) and

$$A \cdot \begin{pmatrix} a\frac{p+q\sqrt{\Delta}}{2} \\ (\frac{b+\sqrt{\Delta}}{2}) \cdot (\frac{p+q\sqrt{\Delta}}{2}) \end{pmatrix} = \begin{pmatrix} \pm t'a' \\ t' \cdot \frac{b'+\sqrt{\Delta}}{2} \end{pmatrix}. \qquad (*)$$

Therefore, $xa\frac{p+q\sqrt{\Delta}}{2} + y\frac{(bp+q\Delta)+(p+bq)\sqrt{\Delta}}{4} = \pm t'a'$ which implies that $xa\frac{p}{2}+y\frac{(bp+q\Delta)}{4} = \pm t'a'$ and $xa\frac{q}{2}+y\frac{p+bq}{4} = 0$.Hence $2xaq = -y(p+bq)$. Let e be the positive g.c.d. of $2a$ and $p+bq$. Then $x\frac{2aq}{e} = -y\frac{p+bq}{e}$ so $\frac{2aq}{e}$ divides y and $y = \frac{2aq}{e}.r$ for some integer r. Then $x = -r\frac{p+bq}{e}$. Since $(x,y) = 1$ we get $r = \pm 1$. A simple calculation now shows that $xa\frac{p}{2} + y\frac{(bp+q\Delta)}{4} = \pm t'a' = -\frac{2ar}{e}\alpha\bar{\alpha}$. Hence, keeping in view the various signs, we get $t'a' = \frac{2a}{e}\alpha\bar{\alpha}$. Furthermore, since $xw - yz = 1$, substituting the values of x and y given above we get $w(p+bq) + 2aqz = -re$. We further get from $(*)$ that $2a\frac{p+q\sqrt{\Delta}}{2} + w(\frac{b+\sqrt{\Delta}}{2}).(\frac{p+q\sqrt{\Delta}}{2}) = t'.\frac{b'+\sqrt{\Delta}}{2}$ which implies that $2zap + w(bp+q\Delta) = 2t'b'$ and $2zaq + w(p+bq) = 2t'$,i.e., $-re = 2t'$. Hence $2zap + w(bp+q\Delta) = -reb'$. It is now easy to check that $A^t.(aX^2 + bXY + cY^2) = a'X^2 + b'XY + c'Y^2$. For instance, the coefficient of X^2, if we replace X by $xX + zY$ and Y by $yX + wY$ in the expression $aX^2 + bXY + cY^2$, is $ax^2 + bxy + cy^2$. Substituting $x = -r\frac{p+bq}{e}$ and $y = \frac{2aq}{e}.r$ and using the fact that $t'a' = \frac{2a}{e}\alpha\bar{\alpha}$ and $2t' = -re$ we get $ax^2+bxy+cy^2 = a'$. Similarly, the coefficient of XY on the required transformation is $2axz + bxw + byz + 2cyw$ which on substitution is just b'. Therefore $A^t.(aX^2+bXY+cY^2) = a'X^2+b'XY+c'Y^2$ and ϵ is injective.

This is a beautiful example in mathematics where two apparently unrelated objects turn out to be equal. Maybe the reader can discover some more.

Suggested Reading

[1] D Flath. *Introduction to Number Theory*. John Wiley and Sons. New York, 1989.

[2] J P Serre. *A Course in Arithmetic*. Narosa Publishing House. New Delhi, 1979.

[3] H M Stark. The complete determination of the complex quadratic fields of Class number one. *Michigan Math J.* 14. 1–27,1967.

[4] Algebraic Number Theory. *Mathematical Pamphlets 4*. Tata Institute of Fudnamental Research. Mumbai, 1964.

Address for Correspondence
Rajat Tandon
Department of Mathematics
University of Hyderabad
Central University P.O.
Hyderabad 500 046, India

"Bhabha is a great lover of music, a gifted artist, a brilliant engineer and an outstanding scientist ... He is the modern equivalent of Leonardo da Vinci."

C V Raman

Enzyme Kinetics? Elementary, my dear ...

2. The Analysis and Significance of Kinetic Parameters

Desirazu N Rao

Desirazu N. Rao is at the Department of Biochemistry, Indian Institute of Science, Bangalore. His main research interests are in the areas of protein-DNA interactions using restriction enzymes as model systems, and in DNA methylation.

[1] See *Resonance*, Vol.3, No.6, 1998.

Catalysis is essential to make many critical biochemical reactions proceed at useful rates under physiological conditions. We had earlier discussed in Part 1[1] the basic principles of enzyme catalysis and derived the Michaelis–Menten equation. In this article, the significance of kinetic parameters and analysis of kinetic data will be discussed.

Why should one determine K_m and V_{max} ?

K_m, the Michaelis constant is a dynamic constant expressing the relationship between the actual steady-state concentrations rather than the equilibrium concentrations. *Table I* shows the K_m values of some enzymes. K_m depends on the particular substrate used, pH, temperature and ionic strength. Observed values of K_m for different substrates and different enzymes vary widely; the smaller values are in the region of 10^{-7} M whereas poor substrates can have very high values. Typical values for physiological substrates are generally in the region of 10^{-3} M to 10^{-6} M.

Inspection of Michaelis–Menten equation shows that K_m is equivalent to the substrate concentration that yields half maximal velocity. If $v = V_{max}/2$ then,

$$\frac{V_{max}}{2} = \frac{V_{max}[S]}{K_m + [S]}.$$

Dividing both sides by V_{max} gives:

$$\frac{1}{2} = \frac{[S]}{K_m + [S]}$$

hence, $K_m = [S]$. Very often, K_m is assumed to be equal to the

Enzyme	Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
Carbonic anhydrase	CO_2	8000	600,000	7.5×10^7
Chymotrypsin	Acetyl-L-tryptophanamide	5000	100	2×10^4
Penicillinase	Benzylpenicillin	50	2,000	4×10^7
Lysozyme	Hexa-N-acetyl glucosamine	6	0.5	8.3×10^4
Pyruvate decarboxylase	Pyruvate	400	–	–
	HCO_3^-	1000	–	–
	ATP	60		
*Eco*RV endonuclease	Plasmid DNA containing one *Eco*RV site	0.0005	0.015	3×10^7
*Hha*I DNA methylase	λ DNA *Bst*EII digest	0.06	0.22	3.7×10^6
*Eco*RV endonuclease	Oligonucleotide containing one *Eco*RV site	3.8	0.115	3.0×10^4

Table 1. Kinetic constants of some enzymes.

dissociation constant for the ES complex. This is true only if the rate constant for the formation of products, k_3 is significantly smaller than k_2. Consider a situation where k_2 is much greater than k_3. The dissociation of the ES complex to E and S is much faster than the formation of E and the product. Under these conditions $(k_2>>k_3)$,

$$K_m=k_2/k_3.$$

The dissociation constant of the ES complex (K_{ES}) is

$$K_{ES}=\frac{[E][S]}{[ES]}=k_2/k_1.$$

This means, K_m is equal to the dissociation constant of the ES complex. Under these conditions $(k_2>>k_3)$, K_m is a measure of the strength of the ES complex – a high K_m means weak binding and a low K_m means strong binding. Most often, $k_3>>k_2$, in which case K_m is not directly equivalent to a dissociation constant for ES. In any case K_m is the concentration of substrate at which half the active sites are filled.

The numerical value of K_m is of interest for several reasons,

(a) The K_m establishes an approximate value for the intracellular level of the substrate.

(b) As K_m is a constant for a given enzyme, its numerical value provides a means of comparing enzymes from different organisms or from different tissues of the same organism or from the same tissue at different stages of development.

(c) A ligand-induced change in the effective value of K_m is one way of regulating the activity of an enzyme. By measuring the effects of different compounds on K_m, it is possible to identify physiologically important inhibitors and activators.

The maximum velocity V_{max} is not by itself a very useful comparative parameter because of its dependence on enzyme concentration.

$$V_{max} = k_3 [E_t] .$$

A more useful parameter is turnover number [2] or k_{cat}, which is equivalent to the rate constant k_3, for the breakdown of the *ES* complex to product when the enzyme is fully saturated with substrate. Since $V_{max}=k_3[E_t]$, the term V_{max} reveals the turnover number of an enzyme if the concentration of enzyme $[E_t]$ is known.

[2] The turnover number of an enzyme is the number of substrate molecules converted into product by an enzyme molecule in unit time when the enzyme is fully saturated with substrate.

$$k_{cat}=V_{max}/[E_t].$$

k_{cat} is a first order rate constant and therefore will have units of reciprocal time. The turnover number is a measure of the maximum potential catalytic activity of an enzyme. The reciprocal of the turnover ($1/k_{cat}$) is the time taken for a single round of catalysis to occur when the enzyme is saturated with substrate. Turnover numbers vary widely, the highest value observed is for carbonic anhydrase (*Table 1*). This means that with a turnover number of 600,000/sec, a 10^{-6} M solution of carbonic anhydrase catalyses the formation of 0.6 M H_2CO_3 per second when it is fully saturated with substrate. Each round of catalysis occurs in 1.7μ seconds

($1/k_{cat}$). In other words, each enzyme molecule can hydrate 10^5 molecules of CO_2 per second.

Another important parameter is the ratio of the turnover number to the Michaelis constant. The value k_{cat}/K_m is a second order rate constant for the reaction of enzyme and substrate to form products.

Consider the following reaction:

$$E+S <----> E+P$$

when $[S]>>K_m$, then,

$$v = \frac{-\mathrm{d}[S]}{\mathrm{d}t} = \frac{k_{cat}}{K_m}[E][S].$$

This ratio is referred to as specificity constant for an enzyme. For example, suppose an enzyme can catalyse a reaction with either of the two substrates A and B, then if the value of k_{cat_a}/K_{m_a} is greater than k_{cat_b}/K_{m_b}, then it follows that substrate A will be utilised at a greater rate i.e., the enzyme has a greater specificity for substrate A than for B. Since an enzyme and substrate cannot combine more rapidly than diffusion permits, there is an upper limit on enzyme catalysis. The value of k_{cat}/K_m cannot be greater than about $10^9 s^{-1} M^{-1}$. Some enzymes like carbonic anhydrase have values of k_{cat}/K_m that approach the diffusion limit, indicating extreme efficiency in binding substrate and in converting it to a product. The specificity constant is a useful parameter when one compares a wild-type enzyme with a mutant enzyme. This is all the more important especially when one is studying amino acid replacements at the active sites of enzymes by site-directed mutagenesis [3]. *Therefore, high catalytic activity and high specificity for substrates, can be described kinetically by the constants* k_{cat} *and* k_{cat}/K_m.

[3] Site-directed mutagenesis is a procedure for producing a protein in which specific amino acid residues have been replaced by others at the DNA level.

Graphical Methods to Estimate K_m and V_{max}

As the v versus $[S]$ curve is a hyperbola (*Figure 3* in Part 1), it is difficult to determine V_{max} and K_m. Many students assume that V_{max} can be estimated from a v versus S plot by finding the point at which v 'reaches' its limiting value. The main drawback lies in drawing a curve that flattens out too abruptly and then drawing an asymptote too close to the curve. An asymptote is a straight line approached by a given curve, as one of the variables in the equation of the curve approaches infinity. Enzyme kinetic data may be treated graphically in the same manner as that of ligand binding. Over the years many graphical and computational methods have been developed in an attempt to improve the accuracy with which the kinetic constants may be determined. The most widely used graphical method for the determination of K_m and V_{max} has been the Lineweaver–Burk plot, named after the two scientists who devised the plot. This is a double reciprocal plot, derived by taking reciprocals of both sides of the Michaelis–Menten equation (equation 4) to give

$$\frac{1}{v} = \frac{1}{[S]}\frac{K_m}{Vmax} + \frac{1}{Vmax} \quad (\text{cf} : y = mx + c)\,.$$

Hence, when $1/v$ is plotted against $1/[S]$ a straight line is obtained with a slope of K_m/V_{max} and an intercept of $1/V_{max}$ on the $1/v$ axis (*Figure 1*).

This plot does suffer from a very serious disadvantage that is often ignored. By taking reciprocals, greater significance is

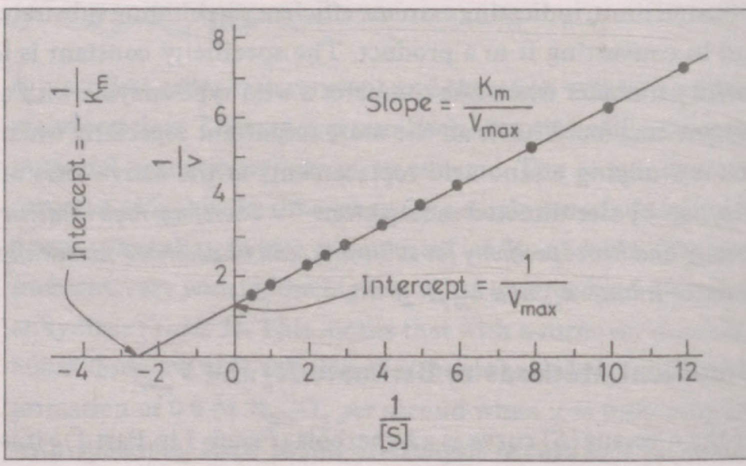


Figure 1. Double-reciprocal (1/v versus 1/[S]) Lineweaver–Burk plot.

placed on the rates obtained at low substrate concentrations and it is these that are most likely to be subject to greatest experimental error. The capacity of the double-reciprocal plot to 'launder' poor data i.e., to provide a visual effect of scatter (but not in reality) probably accounts for its extraordinary popularity with biochemists. It is therefore not recommended as a method for estimating K_m and V_{max}. However, by using suitable weightage, the problems with the double-reciprocal plot can be overcome but often a best fit line appears to the eye to fit badly. The Lineweaver–Burk plot is not the only linear transformation of the basic velocity equation. Indeed, under some circumstances one of the other linear plots may be more suitable or may yield more reliable estimates of the kinetic constants (see *Suggested Reading*).

A plot for obtaining the kinetic constants that has the advantage of simplicity as well as being statistically acceptable is the direct linear plot (see *Suggested Reading*). This method treats V_{max} and K_m as variables and v and $[S]$ as experimentally determined constants according to the rearranged Michaelis–Menten equation:

$$V_{max} /v - K_m/[S] = 1.$$

A plot is constructed with the x-axis labelled K_m and the y-axis labelled V_{max}. For each experimental result, a straight line is drawn from a value of $-[S]$ on the x-axis to the value v, on the y-axis. This line is extended into the first quadrant of the graph and represents a series of values of V_{max}/K_m that would produce a rate v_1 at substrate concentration $[S_1]$. Similarly, lines are drawn for a set of data $[S_2, S_3, \ldots, S_n]$, $v_2, v_3, \ldots, v_n$. The lines should all intersect at a single point with coordinates of K_m and V_{max} which represent the only values of K_m and V_{max} that satisfy all of the sets of data (*Figure 2*). In practice, there is likely to be more than one intersecting point (*Figure 2*, inset). When there are multiple intersections, the best estimates of the true values of K_m and V_{max} are determined by the coordinates of the *median* intersection point. The most obvious advantage of this plot is that the original form of it requires no calculations at all. This allows it to

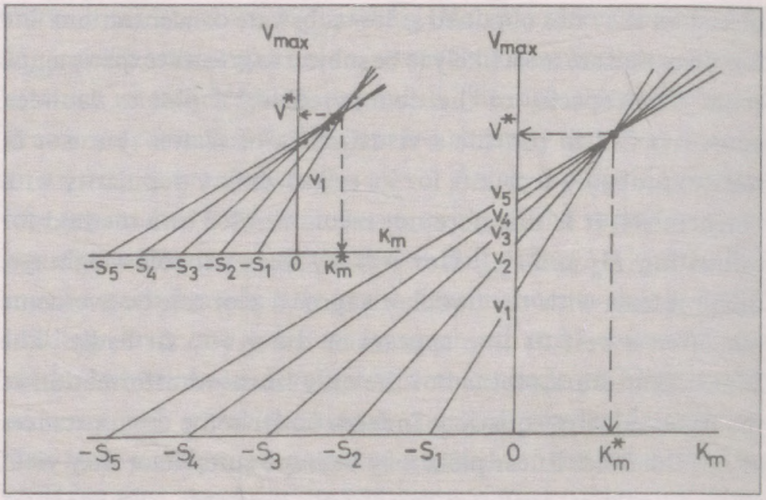


Figure 2. Direct linear plot of V_{max} against K_m. Each line represents one data point and is drawn with an intercept of –S on the x-axis. In an idealised situation, all the lines intersect at a unique point whose coordinates yield the values of K_m and V_{max} that fit the data. Inset shows more realistic data with experimental errors causing the unique point to breakdown into a family of points. The best estimate of K_m and V_{max} can be taken as the medians (middle values) of the two series.

be used very easily in the laboratory while the experiments are proceeding, so that one has an immediate visual idea of the likely parameter values and of design needed for defining them accurately. These features make it more suitable for use in the laboratory for actual analysis of data than for presenting the results in the literature.

Suggested Reading

[1] I H Segel. Enzymes in *Biochemical Calculations*. John Wiley and Sons Publishers. 208–237,1976.

[2] A C Bowden. Introduction to Enzyme Kinetics in *Fundamentals of Enzyme Kinetics*.Portland Press. U.K. Publishers. 19–34, 1995.

[3] Smith and Wood. Enzyme Kinetics in *Biological Molecules*. Chapman and Hall Publishers. 83–94, 1995.

[4] L Stryer. Enzymes: Basic Concepts and Kinetics in *Biochemistry*. W.H. Freeman and Company Publications. 181–195, 1995.

Address for Correspondence
Desirazu N Rao
Department of Biochemistry
Indian Institute of Science
Bangalore 560 012, India
email:
dnrao@biochem.iisc.ernet.in

How do Plants Absorb Nutrients from the Soil?

Study of Nutrient Uptake

G Sivakumar Swamy

The study of nutrient uptake by plant roots has been a fascinating subject both from the academic viewpoint and also its application in crop productivity. This article provides a very brief account of our current understanding of this phenomenon which requires an interdisciplinary approach. Plant cell simulates an electrochemical battery cell, and also represents a complex electronic circuit.

G Sivakumar Swamy obtained his Ph D from the University of Windsor, Canada, in 1980, working on mitochondrial and chloroplast tRNA. He is now teaching plant physiology in the Department of Botany, Karnatak University, Dharwad. His current research interests are in membrane lipids, and membrane-bound enzymes in plants.

Introduction

Animals, including man, require food in the form of carbohydrates, proteins, vitamins, etc., which in turn are provided either directly or indirectly by plants. Then, how do plants obtain their food? Plants have the unique ability to synthesise their own food utilising solar energy and the inorganic elements available in their surroundings. They obtain their carbon, hydrogen, and oxygen from water and from the atmospheric CO_2 and O_2. The soil is the source of other inorganic nutrient elements which are normally available as ions such as NO_3^-, $H_2PO_4^-$, SO_4^{--}, K^+, Ca^{2+}, Mg^{2+}, Fe^{3+}, etc. Like all other living organisms, plants have the ability to maintain an internal environment with a composition different from that of their surroundings. The internal environment (chemical contents) of the plant body remains more or less constant whereas the outside environment is highly variable. It is a fact that the concentration of the nutrient elements and other molecules are far greater inside the plant body compared to the outside environment. Still the plant roots are able to absorb nutrients against a high concentration gradient which would have resulted otherwise due to the natural physical forces operating in the system. This property of holding a solution of higher concentration inside the cell against a dilute solution

Box 1. Fluid Mosaic Structure

The widely accepted model of membrane structure is the 'fluid mosaic' model. This model includes the 'fluid' lipid bilayer interspersed with both intrinsic and extrinsic proteins, some of which are structural proteins and others are involved in various membrane functions (*Figure 1*). The interior of the membrane consists of the hydrocarbon chains of fatty acid moieties of the lipids arranged in the form of a bilayer, and as a result it is strongly non-polar. The nonpolar membrane interior acts as a strong barrier for the passage of ions and polar molecules including water. The membrane surface includes the polar groups of the phospholipids and glycolipids, and it is also associated with the polar domains of the membrane proteins. The non-polar domains of the proteins establish hydrophobic interactions with the fatty acid moieties of the lipid bilayer.

outside is the unique characteristic feature of the cell membrane that surrounds the cytoplasm of the cell. It is evident that the plant cells have to overcome the physical forces such as diffusion in order to maintain a higher solute concentration and also to absorb nutrient elements which demands a massive investment of energy by the plant cells.

The Cell Membrane is the Site of Nutrient Transport

An appreciation of the structure of the cell membrane is essential for understanding the mechanism of nutrient uptake by plant roots (see *Box 1*).

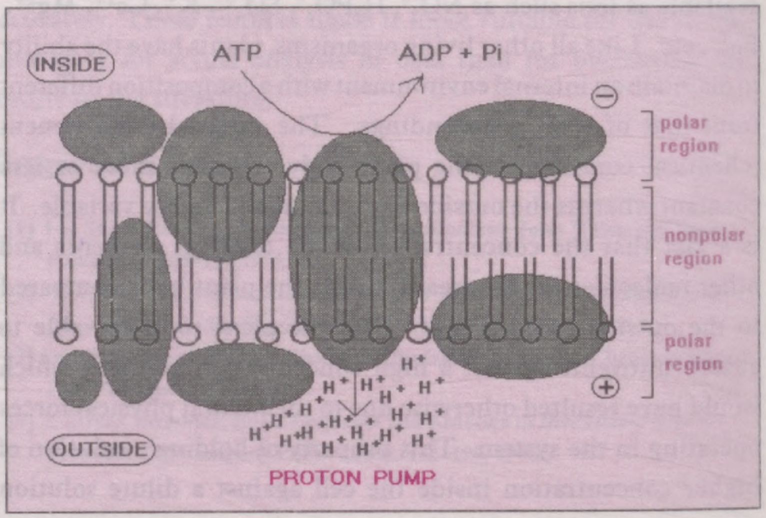


Figure 1. Cross sectional view of the cell membrane: The fluid mosaic structure. Lipid bilayer and proteins form the central non-polar and the peripheral polar regions. H^+-ATPase (proton pump) pumping protons is also shown in the figure. The + and - signs inside the small circles represent the positive and negative transmembrane potentials outside and inside the cell, respectively.

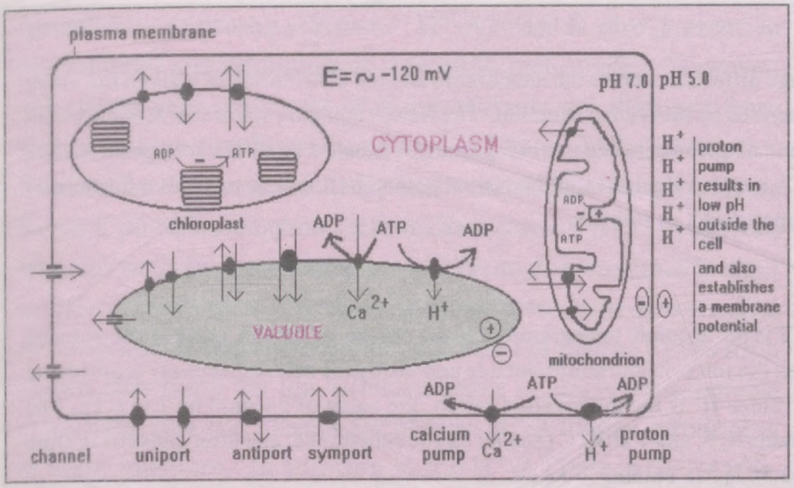


Figure 2. Diagram of a typical plant cell (cell wall not shown) showing the transport of ions and molecules across the membranes. ATP synthesis is driven by the proton motive force generated by electron transport in mitochondrion and chloroplast. On the other hand, hydrolysis of ATP mediated by H^+-ATPase generates proton motive forces across plasma and vacuolar membranes resulting in the generation of transmembrane potential (shown as + and - signs within small circles). Ca^{2+} pumping by Ca^{2+}-ATPase is also shown. Transport processes by uniport, symport, and antiport (see text for details) are also represented. (Note: Pi released/ utilised not shown).

This structural design of the membrane offers scope for the transport of nutrient elements only through integral proteins of the membranes. These transport proteins are very similar to enzymes in their specificity in recognising the molecules and ions for transport. The driving force for the transport process is provided by the membrane-bound ATP hydrolysing enzyme, H^+-ATPase (*Figures 1, 2*).

H^+-ATPase Transports Protons and Generates Transmembrane Potential

Whenever an ion moves into or out of a cell unbalanced by a counter ion of opposite charge, it creates a voltage difference across the membrane called electrogenic pump. The functioning of the electrogenic pump is an energy consuming process and it provides the driving force for the ion transport across membranes by generating a transmembrane potential. In animals, the $(Na^+ + K^+)$-ATPase serves as the electrogenic pump while in plants H^+-ATPase serves this purpose. This plant enzyme, being located in the plasma membrane, generates a pH gradient across the membrane and also establishes a transmembrane potential by the vectorial pumping of protons from the cytoplasm to the exterior (*Figures 1, 2*). Under ideal conditions (when no other ions are interfering), a transmembrane potential of about –120 mV is generated inside the cell as the result of a pH difference of about

Box 2

A relationship between voltage difference across the membranes and the distribution of a given ion under equilibrium conditions is described by the Nernst equation. The Nernst equation states that at equilibrium the differences in concentration of an ion between two compartments is balanced by the voltage difference between the compartments. A simple calculation of the Nernst potential E can be made as a function of H^+ concentration in two compartments as follows:

$$E = (RT/zF)\ \ln\ (C^0/C^i)\ ,$$

where R is the gas constant, T is the absolute temperature, z is the charge of the ion, and F is the Faraday constant. C^0 and C^i represent the ion concentrations outside and inside the cell respectively (considered here as two compartments). Since H^+ is the monovalent cation, the value of z would be 1 ($z = 1$). The numerical values of the constants R, F, and T at 30°C (303 K) can be substituted, and after converting from natural logarithm to $\log_{10}$ (x2.303), we obtain:

$$E = 60\ \log\ (C^0/C^i).$$

Suppose the H^+ concentration across the membrane is 10^{-5} M (pH 5.0) and 10^{-7} M (pH 7.0) as shown in *Figure 2*, then, $E = 60\ \log 100$, and therefore, the membrane potential inside the cell is –120 mV.

2.0 units across the membrane (see *Box 2*). However, under natural conditions, membrane potentials of more than –120 mV are commonly observed in plant cells due to interaction with other ions in the outside medium as well as in the cytoplasm.

The membrane proteins that are involved in the transport process may be classified as: (i) pumps, (ii) carriers and (iii) channels (*Figure 2*). The proton-pump ATPase (H^+-ATPase) acts as a primary transporter by pumping protons out of the cell. This process creates a pH and electrical potential difference across the membrane. It is estimated that the H^+-ATPase alone utilises 25–50% of cellular ATP which indicates the amount of energy invested by the plants to absorb nutrients, and to maintain a higher solute concentration inside the cell. The plasma membrane H^+-ATPase is composed of a single polypeptide of a molecular mass of about 100 kDa. It transports one proton per molecule of ATP hydrolysed, and has a pH optimum of about 6.5 for catalysis. The plasma membrane H^+-ATPase protein stretches across the thickness of the membrane (transmembrane protein) and has ten segments of the polypeptide extending across the membrane

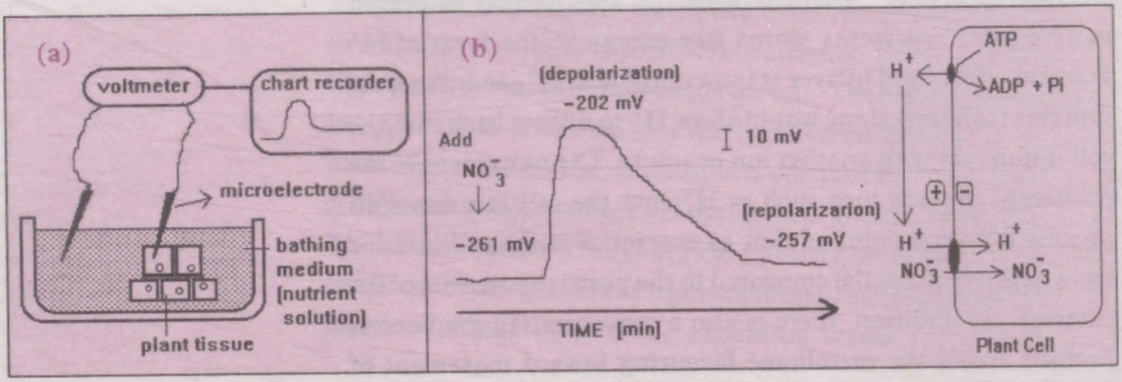
(membrane-spanning regions). H^+-ATPase is also present in the vacuolar and endoplasmic reticulum membranes. However, the amino acid sequences of those enzymes are different from that of plasma membrane-bound H^+-ATPase. As far as the reaction mechanism is concerned, plasma membrane-bound enzyme forms a phospho-enzyme intermediate during ATP hydrolysis whereas such an intermediate is not formed in enzymes located in other membranes. Further, there are also tissue-specific and developmental stage-specific H^+-ATPases in plants. It has been shown that there are ten different genes in *Arabidopsis thaliana* expressing just the plasma membrane H^+-ATPases, producing as many enzymes. In addition to the H^+-ATPase, a Ca^{2+} translocating ATPase has also been identified in the plant cell membranes. The Ca^{2+}-ATPase is involved in pumping out Ca^{2+} from the cytoplasm into the cell compartments or to the apoplast (outside the cell). This pump is involved in the control of cytoplasmic Ca^{2+} level which in turn regulates the process of signal transduction involving the calcium binding protein, calmodulin.

Carrier Proteins and Ionic Channels Serve as Conduit for Nutrient Transport

The transport of nutrient ions across the membrane can be explained by invoking the chemiosmotic principle which was originally put forth by the Nobel Laureate, Peter Mitchell. When protons are extruded from the cell electrogenically, both a transmembrane potential and a pH gradient are created at the expense of cellular energy (released by ATP hydrolysis). This electrochemical H^+ gradient which is also termed as proton motive force represents stored free energy in the form of H^+ gradient. The lipid bilayer is impermeable to H^+, and transport proteins (carriers) alone would allow H^+ to diffuse back into the cell if it moves with another ion or solute. The movement of the positively charged ions such as H^+ into the cell is a downhill process from the point of view of energetics as the cell interior has a negative potential compared to the positive potential of the exterior. In addition, there is also a concentration gradient of protons across the membrane favouring inward movement of protons. If this downhill process of H^+ transport is linked to the cotransport (symport) of an oppositely charged ion into the cell, the cell can acquire a negatively charged ion against its concentration gradient (even though that particular anion is in excess amount inside the cell compared to the external soil solution). Since the free energy change (ΔG) of H^+ transfer from the exterior into the cell interior is strongly negative this process can also drive the transport of an anion or a neutral molecule against their concentration gradient.

An example of symport mechanism of transport can be cited in the uptake of NO_3^- by the plant cells. The membrane potential of the intact cell can be determined by introducing a microelectrode into the cell as shown in *Figure 3a*. As represented in *Figure 3b* there occurs an initial depolarisation of the membrane if NO_3^- is added to the bathing medium of the cell. These results would imply that NO_3^- is transported into the cell as H^+/NO_3^- symport mediated by an NO_3^- carrier protein. The symport of these two ions into the cell would decrease the H^+ concentration outside the cell, and this is expressed in the reduction of transmembrane potential (depolarisation) as shown in the figure. The movement of H^+ into the cell along with NO_3^- would result in the net transport of NO_3^- into the cell (*Figure 3b*). The depolarisation of the membrane would stimulate H^+-ATPase and this phenomenon results in the repolarisation (increase in the transmembrane potential) followed by the earlier transient depolarisation. Since the cell interior exhibits negative potential (which occurs at the expense of cellular energy) the uptake of

Figure 3. (a): Diagrammatic representation of the experimental set-up for measuring membrane potential. A glass microelectrode is introduced into the cell bathed in the nutrient medium. Change in the potential by the addition of a nutrient into medium can be continuously monitored by the voltmeter attached to a recorder. (b): Diagram showing the recording of the membrane potential in the experimental set-up shown in (a). Addition of NO_3^- into the nutrient medium bathing the plant tissue results in transient depolarisation followed by repolarisation. The diagram of the plant cell on the right side explains the symport mechanism of NO_3^- transport across the plasma membrane.



positively charged nutrient ions (cations) into the cell can take place by an uniport mechanism without the cotransport of protons. Similarly, anions can be extruded through specific carriers from the cell interior to the exterior since the latter has a positive potential. Two different anions can be exchanged across the membrane by an antiport mechanism resulting in the net uptake of one of the anions at the expense of the efflux of the other.

In addition to the carrier proteins, there are channels for ion transport which are also highly specific in recognising and transporting the designated ions across the membrane (*Figure 2*). The channels are the openings in the membrane, and are surrounded by proteins which would specifically allow a particular ion to pass through. The channels would invariably carry out uniport movement of ions driven by the electrochemical gradient. A number of channels have been identified in plant cells for the transport of ions such as K^+, Ca^{2+}, Cl^-, etc., and also for the transport of water. The water transporting channels are known as aquaporins. These channels possess the unique characteristic feature of opening and closing at definite time intervals, and this phenomenon is governed by the thermal motion. In addition, the opening of channels is also governed by the transmembrane potential (voltage-gated channels), and also influenced by regulatory molecules (ligand-gated channels). There are also 'stretch-activated' channels in plants which are regulated by the cell turgor. From the functional point of view, both uniport carriers and channels are similar except that the rate of transport through the channels is several fold higher compared to that of carriers.

Transfer of Genes Encoding Transport Proteins has an Enormous Economic Potential

So far, the transporter (carrier) genes for NO_3^-, $H_2PO_4^-$ SO_4^{--}, and K^+ transport have been cloned from the plants. This appears to have paved the way for transferring the genes encoding transporter proteins into crop plants. It is important to point out that high yielding crop plants require higher levels of fertilisers.

Suggested Reading

[1] S R Assman and L L Haubrick. Transport proteins of the plant plasma membrane. *Curr. Opinion Cell Biol.* 8. 458–467, 1996.

[2] H Logan, M Basset, A-A Very and H Sentenac. Plasma membrane transport systems in higher plants: From black boxes to molecular physiology. *Physiol. Plant.* 100. 1–15, 1997.

[2] L Taiz and E Zeiger. *Plant Physiology*. Benjamin/ Cummings Publishing Company, 1991.

Box 3. Genes Encoding Transport Proteins have been Cloned.

A few genes coding for transport proteins have been cloned and made to express *in vivo*. Different strategies have been adopted to clone these genes, and the cloning of NO_3^- transporter can be considered here as an example for study. The commonly used herbicide, chlorate causes yellowing of leaves culminating in defoliation and death of the weeds and other herbs. A mutant (CHL1) of *Arabidopsis thaliana* was obtained which lacked the sensitivity to the chlorate herbicide. This mutant was unable to take up chlorate into the cell, and therefore, the plants became resistant to the herbicide. Incidentally, chlorate (ClO_3^-) is a structural analog of NO_3^-, and therefore, is transported into the cell mediated by the NO_3^- carrier. The CHL1 mutant is capable of growing in the presence of chlorate because it is defective in chlorate uptake, and consequently, NO_3^- uptake is also inhibited.

However, the CHL1 mutants could be grown in the presence of nitrogen in some form other than NO_3^- (such as ammonium salt). An insertional mutant of chlorate uptake (invariably defective in NO_3^- transport) was also obtained by using T-DNA of the bacterium, *Agrobacterium tumefaciens*. This bacterium consists of a plasmid known as Ti plasmid which has a segment in it referred to as T-DNA. When the bacterium infects the plant the T-DNA segment of Ti plasmid is transferred to the host chromosome. The T-DNA integrates into the plant DNA and expresses itself along with plant genes. The insertion of T-DNA into the plant chromosome is random, and therefore, there is a chance that it inserts itself into the NO_3^- transporter gene and causes an insertional mutation. When a large number of transgenic plants with T-DNA insertions were screened in the presence of chlorate only a few plants were found to be green and healthy. These plants were later found to be insertional mutants of chlorate (also NO_3^-) transport. A genetic cross between the insertional mutant and the original CHL1 mutant indicated that both the genes cosegregated which indicates that both mutations occurred in the same locus. The DNA from the insertional mutant was isolated and the CHL1 gene was identified by hybridising with labelled T-DNA. This had enabled cloning the flanking regions of the T-DNA which was in fact the NO_3^- carrier gene. This gene was cloned and transcribed *in vitro* to obtain the mRNA. When the mRNA of this gene was microinjected into the oocytes of the toad *Xenopus laevis*, the NO_3^- carrier protein was synthesised in the toad oocytes and exhibited all the characteristics of NO_3^- transporter in the *Xenopus* oocyte including the ability to transport NO_3^- into the cell. Similarly, the gene encoding K^+ transporter was isolated by using a different strategy from wheat roots and made to express in the yeast mutant which lacked the ability to take up K^+.

This may have resulted because efficiency in nutrient uptake was not addressed while breeding the crop plants for higher yield. On the other hand, there are a large number of plants which grow luxuriantly in wild in the soils which are poor in nutrient content. These plants may manage to grow well in such soils possibly because they have efficient nutrient uptake systems preserved in them by natural selection. Therefore, there is a bright prospect for cloning the transporter genes from those plants and transferring them to crop plants so that crop plants would become more efficient in nutrient uptake thereby bringing down their fertiliser requirement.

Address for Correspondence
G Sivakumar Swamy
Department of Botany
Karnatak University
Dharwad 580 003, India.

Game Theory

1. Nash Equilibrium

P G Babu

This article tries to outline what game theory is all about. It illustrates game theory's fundamental solution concept viz., Nash equilibrium, using various examples.

The author is an Assistant Professor at Indira Gandhi Institute of Development Research, Mumbai.

The Genesis

In the late thirties, the mathematician John von Neumann turned his prodigious innovative talents towards economics. This brief encounter of his with the day's economic theory convinced him that it was in need of a new mathematical tool. In the years that followed, he along with Morgenstern went about creating a brand new mathematical tool (anticipated by Borel and before him by Cournot). This new tool was offered to the profession in their now classic book '*Theory of Games and Economic Behavior*'. In this book, they developed two person zero sum games and other cooperative game theoretic concepts. But, soon economists found out that the phenomenon of 'one person's gain is the other person's loss' was too restrictive in many applications. After some initial euphoria, the interest in this new tool died down except for a small hard core group of mathematicians who continued to work on these concepts. Princeton was the epicentre for most of them. Hence, it is not a surprise to see young Nash Jr taking the next giant step in the Fine Hall of Princeton towards what we now know as 'modern non-cooperative game theory'.

Game theory can be viewed as an interactive decision theory. It deals with the situations where people with different (mostly competing) goals try to take into account others' actions in deciding on the optimal course of action. Take for instance chess. When you decide what move to make, you also take into account the likely response of the opponent and your next response, his reply and so on. The fact that your opponent has

Box 1. John Nash Jr.

John F Nash Jr came to Princeton for his graduation in mathematics with a one line recommendation which said "This man is a genius" Within two years of his arrival, when he was just twenty, Nash completed his 21 page doctoral dissertation under Al Tucker; this dissertation lays down much of what we now know as modern game theory, with the famous equilibrium concept (which stands in his name) to analyse n-person, non-zero sum situations as the centre piece. Even before his arrival at Princeton, Nash Jr, as an undergraduate student, wrote a term paper for a course on international economics (his only encounter with economics). This term paper later on became one of the all time great papers in game/economic theory which laid the foundation stone for bargaining theory. Soon afterwards, Nash Jr was struck by schizophrenia and lost his next thirty years to this disorder and recovered miraculously in 1989 to win the Nobel Prize in Economics for the year 1994 along with John Harsanyi and Reinhard Selten.

equal intelligence and self interest enables you to duplicate his reasoning process. For example, consider two companies producing the same product and competing in the same market. Their aim would be to capture as much market share as they can using various ploys such as price cutting, gift schemes, and advertising. Which one of these ploys would be successful given the opponent company's ploy? Game theorists are interested in exploring such situations.

Eternal Dilemma

The best way to learn game theory is through playing games. Let us begin with one of the most popular games, viz., *prisoners' dilemma*. This game has been attributed to Al Tucker of Princeton University. In this game there are two thieves who have been caught by the police and brought before the magistrate. As there is only circumstantial evidence to their crime, the magistrate comes up with the following clever scheme. He locks them up in separate cells in such a way that they cannot communicate with each other. If they both plead guilty, they get 10 years imprisonment each. On the other hand, if both plead not guilty they get away with 1 year prison terms. But, if one pleads guilty

		PRISONER 1	
		CONFESS	DON'T CONFESS
PRISONER 2	CONFESS	-10, -10	-0.5, -15
	DON'T CONFESS	-15, -0.5	-1, -1

Figure 1. Prisoners' dilemma.

and the other not guilty, then the one who pleads guilty gets 0.5 years and the other gets 15 years imprisonment. The question now is: did the magistrate do the right thing by offering this scheme to the prisoners?

The actual game structure of this decision situation is given in *Figure 1*. This structure is known as the normal form game where the sequences of moves and countermoves (in other words, the temporal structure) are suppressed. Players are assumed to move simultaneously and choose one of the two actions: to plead guilty or not guilty. The actions of the two players result in an outcome. An outcome is generally marked by the payoffs. These payoffs are given by the numbers appearing in each cell corresponding to their actions. The first element stands for the row player's payoff and the second corresponds to that of the column player. Players are assumed to be rational economic agents who are interested in maximising their payoffs. The following features of the game are common knowledge ('everyone knows it, everyone knows that everyone knows it, ...'): the rationality of the players, the action choices, and the payoffs.

Players are assumed to be rational economic agents who are interested in maximising their payoffs.

Let us see if the magistrate did the right thing or not. Look at the row player. His best payoff is 1 year prison term but to get that he needs to plead 'not guilty'. But, his co-conspirator now has the incentive to plead guilty as she can get away with 6 months imprisonment. Given this fact along with the knowledge that she is a rational player leads the row player to believe that if he were in her position he would choose to plead guilty. With this reasoning at the back of his mind, he prefers to plead guilty. The column player also reasons along the same lines and opts to plead guilty. Note that pleading guilty is the only possible solution in this game. Hence, both end up pleading guilty vindicating the magistrate's scheme.

This game brings about the contradiction between the individually rational outcome (both pleading guilty) and what is collectively good for all (both pleading not guilty). It also shows that the resulting solution to the game could be Pareto inefficient (See *Box 2*).

Box 2. Pareto Efficiency

Pareto efficient solution is one where the players cannot in any way improve their current payoffs through a different action choice without reducing others' payoffs. In the above game, the outcome where both the prisoners are 'pleading guilty' can be improved upon by both deciding to plead 'not guilty' without hurting the other player's payoff. Hence, we conclude that the unique Nash equilibrium of the game is Pareto inefficient.

Note that conjectures about the opponent's play has no role in picking the final solution in this game. The equilibrium concept we have used to solve the game is what is known as the Nash equilibrium: choose the best action from among your action set given that the opponent will choose her best action. Of course, we assume that the action sets, payoffs and the game structure are common knowledge. This game can be used to study various issues such as: two firms competing to sell the same product (say, a toothpaste), two nations erecting trade barriers and the exploitation of common property resources [1] such as fisheries.

[1] These are the resources for which no one has exclusive property rights and hence none have any incentive to protect the resource or optimally use it.

Notice that the key to arriving at the Nash equilibrium outcome in the above game is the individual's ability to duplicate other's reasoning process. But if you do not know the other person's characteristics, his tastes, ability, ideology, etc., then you can not reason what he will do in a given situation. In real world, such lack of knowledge about one's opponent is the norm. For a long time, game theorists did not know how to formulate such a situation. This proved fatal to the applicability of game theory to real world problems. In this situation, John Harsanyi provided a breakthrough. In a three part paper written in early 1960s, he showed how one can formulate games where people do not have much information about each other. It would be apt to say that Harsanyi was the one who resurrected game theory's applicability. This formulation of Harsanyi is however beyond the scope of this paper.

Box 3. John Harsanyi

Harsanyi, born in a pharmaceutical family in Hungary (1920) proved his early mathematical prowess by winning the first prize in high school mathematics in the whole of Hungary. Later on he vacillated between various subjects including botany and finally went on to earn a doctorate in philosophy from the University of Budapest. In the aftermath of the second world war, Harsanyi migrated to Australia where he worked as a factory worker and later as a clerk for the association of the coal mining industry. To improve his career opportunities, he learnt economics during his spare time and wrote a few papers in international economics to win a job as a lecturer at the University of Queensland. From there he went to Stanford University in the USA on a Rockefeller fellowship and wrote his second doctoral thesis, this time in economics, in 1959. After a few years in Australia, he went back to the USA to join the faculty of University of California, Berkeley from where he retired in 1990. Harsanyi's academic career offers a complete contrast to the fairy tale career of John Nash Jr along with whom he shared a Nobel Prize in 1994.

Rational Pigs

Let us move on to the next game situation (*Figure 2*). This is a story of two rational pigs. Though both the pigs are rational, one is weak (called subordinate pig) while the other one is strong (christened as dominant pig). These two pigs are put in a cage. There is a lever at one end of this cage which when pressed delivers (for quantification) exactly six grains of maize at the other end. Now, if the pigs want food, they need to learn to press the lever and run to the other end. If the subordinate pig presses the lever, the dominant pig can eat all the six grains. On the contrary, if the dominant pig is the one to press the lever, the subordinate pig can eat up five grains

Figure 2. The rational pigs.

		DOMINANT PIG	
		PRESS	DON'T PRESS
SUBORDINATE PIG	PRESS	1.5, 3.5	-0.5, 6
	DON'T PRESS	5, 0.5	0, 0

before the dominant pig arrives. It is observed that once the dominant pig reaches the grain dispenser, it can get everything that is remaining. In the unlikely event of both the pigs pulling the lever together, the subordinate pig (given its lean constitution) can run faster and eat two grains. Pressing the lever and running to the grain dispenser involves expenditure of some calories and for the sake of quantification, we assume that it can be measured in grain units, viz., 0.5 grains. Who will learn to press the lever?

The above normal form game corresponds to the decision problem facing our two rational pigs. For the subordinate pig, whatever the dominant pig does, not pressing the lever is the best strategy (in the sense that its payoff is higher than the other strategy 'pressing the lever'). Now, what would the dominant pig do? It has to form conjectures about the subordinate pig's behaviour. If it believes that the subordinate pig would press the lever, then, the best strategy available is to wait near the dispenser. But, putting itself in the subordinate pig's shoes, the dominant pig would realise that the subordinate pig will not press the lever. Given this logical conclusion, it is optimal for the dominant pig to press the lever and get 0.5 grains net compared to starvation. Hence, the dominant pig will learn to press the lever. At times, weakness could be strength[2].

[2] Remember what the first husband of Mia Farrow had to say about her in Woody Allen's movie 'Husbands and Wives'.

In this game there is no conflict between individual rationality and collective rationality unlike the prisoners' dilemma game.

Setting the Shop

Let us move on to the third story. Here we face two pav bhaji vendors on Juhu beach trying to find proper locations for their shops. They have to charge the same price to survive but, they can locate anywhere they want on a straight line beach. Their customers are spread out all through the beach and are in general averse to walking. Hence, they would prefer to buy from a shop which is closest to them. Where should the pav bhaji vendors locate their shops?

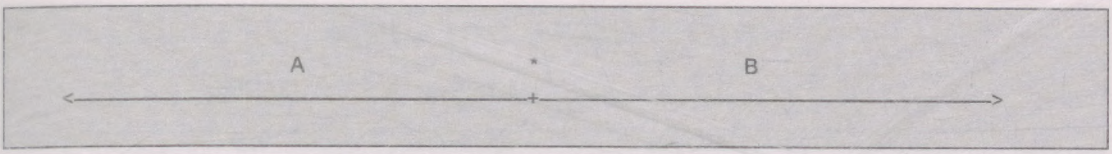


Figure 3. Location game.

Unlike our two earlier games, here players are forced to form dynamic conjectures. Given the nature of the problem, all the customers to the left of, say, pav bhaji vendor A will buy from him and those to the right of B would buy from her. Of those in between them, half will go to A and the rest to B. Now, A reasons as follows: if I can shift my shop one yard closer to B then I can get that many extra customers; but, B being a rational seller will reason the same way and hence she would also move a yard closer to my location. Given that, I need to go two yards closer to B and she would also do the same reasoning and move two yards closer.Hence, I need to go three yards closer... . Ultimately, this logic will stop when A decides to locate the shop right next to that of B and B also does the same thing and it is not difficult to figure out that this location will be exactly in the middle of the beach.

Unlike the previous games, here the logic is fairly subtle. In the previous games, either both or at least one player had a unique best response to whatever the other player did. In this location game, both the players have to simultaneously form conjectures about their rival's choice.

There are many real world situations which bear out the logic inherent in this game. You would have noticed that petrol pumps are located close to each other. The same is true for shops selling foreign goods: witness the Burma Bazaars of Madras, Bangalore and Trichy. Also, next time when you fly notice how different airlines schedule their flights close to each other on a given route.

In this location game, both the players have to simultaneously form conjectures about their rival's choice.

One too Many

You might have noticed that in some sense the previous location game involved dynamic reasoning. In all the games we have seen

		WIFE	
		MOVIE	SAREE
HUSBAND	MOVIE	1, 1	0, 0
	SAREE	0, 0	1, 1

Figure 4. Battle of the sexes.

so far, there have been more or less unique solutions. Could it be that Nash equilibrium always yields an unique solution? Alas! How I wish it were true! But, unfortunately the answer to that question is 'no'. See for example the following game called 'battle of the sexes'. There is a husband and his wife who have to decide where to go for the evening outing. The husband wants to go to a movie while the wife wants to go to the saree shop. Being newly married, wherever they go, they go together. Their problem can be formalised as follows: A careful perusal of this game will tell us that there are two Nash equilibria (to be precise, two pure strategy Nash equilibria). One is when both the husband and wife coordinate on movies and the other when they decide to go to a saree shop. We cannot a priori predict which one of these two solutions will occur. This example alerts us to the problem of multiplicity of Nash equilibria which is a rule rather than an exception in real life decision situations.

So far we have not explicitly introduced the sequential (or temporal) nature of moves by players. In the next part we will address this in a more realistic structure.

Acknowledgements

This article is based in part on a lecture given at the Centre for Theoretical Studies, Indian Institute of Science, Bangalore as part of a lecture series on 'learning and evolution'. I am grateful to Vivek Borkar and Sanjay Jain for this opportunity. I thank the editor-in-charge of the paper for valuable comments. This is an expository article; no originality, other than that of the organisation of material, is claimed.

Suggested Reading

[1] Nash J F. Equilibrium Points in n-person Games. *Proceedings of the National Academy of Sciences (USA)* 36. 48–49, 1950.

[2] Nash J F. Non-cooperative Games, *Annals of Mathematics*. 54. 286–295, 1951.

[3] von Neumann J and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press. Princeton. NJ, 1944.

Address for Correspondence
P G Babu
Indira Gandhi Institute of Development Research
Gen. A K Vaidya Marg
Goregaon(East)
Mumbai 400 065, India.

Machine Translation

A Gentle Introduction

Durgesh D Rao

Machine translation is the study of designing systems that translate from one human language into another. This is a hard problem, since processing natural language requires work at several levels, and complexities and ambiguities arise at each of those levels. Some pragmatic approaches can be used to tackle these issues, leading to extremely useful systems. This article introduces the main concepts, issues and techniques involved in machine translation, and looks at some applications.

Durgesh D Rao is a senior staff scientist with the knowledge based computing systems (KBCS) division at the National Centre for Software Technology, Mumbai. His areas of research interest are natural language processing, information retrieval and machine translation.

Introduction

Imagine this scenario. You dial your colleague in Tokyo. You do not speak Japanese, and he does not speak English. Yet, you are able to converse! You speak into the phone in English, which automatically gets translated into Japanese for him. He replies in Japanese, and you hear it in English. Such a scenario is not yet a complete reality[1], but it is a possibility that has excited and engaged researchers in a field known as *machine translation (MT)* for a number of years.

Machine translation is an important sub-discipline of the wider field of *artificial intelligence (AI)*. AI (among other things) deals with getting machines to exhibit intelligent behaviour. As you might imagine, both AI and MT are interesting and challenging fields. In this article, we will look at the important concepts, issues and techniques in MT.

A machine translation system essentially takes a text[2] in one language (called the *source language*), and translates it into another language (called the *target language*). The source and target languages are natural languages such as English and Hindi, as opposed to man-made languages such as C or SQL.

[1] The 'translating telephone' technology is expected to mature in the 2010s. The Janus project at the Carnegie Mellon University is an example of current efforts in speech-to-speech translation (see *Box 1*).

[2] In this article, we assume text-to-text translation. Some applications, such as the translating phone, may need speech input or speech output, or both, but those can be handled by speech recognition and synthesis modules respectively, with a text-to-text translation system in between.

Box 1. Research MT System Example: The 'Janus' Translating Phone Project

The Janus project at the Interactive Systems Lab, Carnegie Mellon University, is working on a set of translation projects. One important prototype is the 'Translating Videophone Station'.

This prototype system allows two users to communicate in a given domain via a videoconferencing connection. Each party sees the other conversant, hears his/her original voice and sees/hears translation of what he/she says as subtitles, captions and/or synthetic speech. The situation is cooperative, that is, both users want to understand each other and collaborate via the system to achieve understanding.

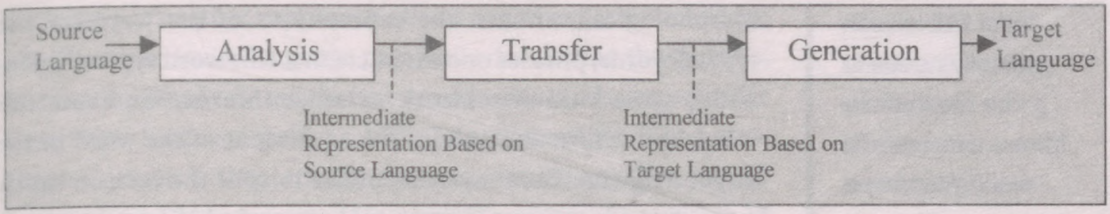
After the record button is activated, the station accepts spoken input and produces a paraphrase of the input sentence first. Once the user has verified that the system properly understood the intended meaning, he/she activates the 'send' button to send a translation of this intended meaning to the other site in the desired language. Various interactive correction mechanisms facilitate quick recovery, should possible processing errors and miscommunications have altered the intended meaning.

Using the technology of this prototype, ISL is also developing prototypes for a portable translation system based on laptops, and simultaneous translation of two speakers in a dialogue.

For more information, visit the ISL homepage at http://www.is.cs.cmu.edu/

Hence an MT system can be said to be doing *natural language processing (NLP)*. In fact as we shall see later, most machine translation applications require some degree of natural language understanding to do the translation.

Machine translation as a discipline dates back to the early nineteen-fifties. The complexity of the problem was originally underestimated, and some early successful demonstrations of experimental systems lead to unrealistic expectations which were hard to fulfil. This led to some skepticism, and funding on MT work almost ceased. In the early eighties, the Japanese Fifth Generation Computing Project revived interest in this work. The current approach to MT is more pragmatic and realistic. It is now widely accepted that *fully automatic, general-purpose, high quality* machine translation is a very difficult problem, but very useful and practical systems can nevertheless be developed by relaxing one or more of these criteria, and several useful systems have been built by doing so, and are in use today. Such systems are being used to translate public announcements, weather bulletins, technical documents, and web pages. Some machine



translation services are starting to become available on the World Wide Web. For example, the web page of the Altavista search engine (http://www.altavista.digital.com) also provides a translation service that can translate simple sentences among a handful of languages.

Components of an MT System

We can divide the machine translation task into two or three main phases – the system has to first *analyse* the source language input to create some internal representation. It then typically manipulates this internal representation to *transfer* it to a form suitable for the target language. Finally, it *generates* the output in the target language.

A typical MT system contains components for analysis, transfer and generation as shown in the diagram. These components incorporate a lot of knowledge about words (*lexical* knowledge), and about the language (*linguistic* knowledge). Such knowledge is stored in one or more *lexicons*, and possibly other sources of linguistic knowledge, such as *grammar*. The user interface is invariably a crucial part of most MT systems. The interface allows users to verify, disambiguate and if necessary correct the output of the system. Another common feature of NLP work is the use of large '*corpora*' (plural for 'corpus'). A *corpus* is a large collection of text which has been appropriately tagged, and is used for acquiring the required lexical and linguistic knowledge.

A *corpus* is a large collection of text which has been appropriately tagged, and is used for acquiring the required lexical and linguistic knowledge.

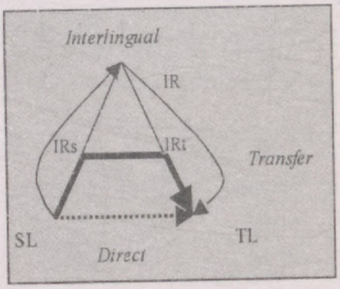
The *lexicon* is an important component of any MT system. A lexicon contains all the relevant information about words and phrases that is required for the various levels of analysis and generation. A typical lexicon entry for a word would contain the following information about the word: the part of speech, the

An MT lexicon typically needs to be much more formal, precise and elaborate than a typical human dictionary.

morphological variants, the expectations of the word (or the typical words, phrases or constructs that this word typically goes with), some kind of semantic or sense information about the word, and information about the equivalent of the word in the target language. Some systems prefer to split the lexicon into a *source* lexicon, a *target* lexicon, and a *transfer* lexicon that maps between the two. The exact format of the lexicons is a matter of engineering design, and would take into account the system designer's policy about issues like how to handle morphological variations, multiple word senses, synonyms and so on. An MT lexicon typically needs to be much more formal, precise and elaborate than a typical human dictionary, since it is meant for mechanical processing, and not for reading by humans. The lexicon plays a central role in modern MT systems.

Approaches to Machine Translation

Based on how closely the internal representation depends on the source and target languages, approaches to MT can be divided into three major classes – *direct*, *transfer-based* and *inter-lingual*, as illustrated in the diagram.



A *direct* MT system tries to directly map the source language to the target language, and is therefore highly dependent on both the source and target languages. A *transfer-based* approach first converts the source language into an internal representation (IRs) which is dependent on the source, but not the target language. The system then transforms IRs into a form (IRt) which is independent of the source language and depends only on the target language, and finally generates the target language output from IRt. The *inter-lingual* approach converts the input into a single internal representation (IR) that is independent of both source and target languages, and then converts from this into the output.

The difference between these approaches becomes clear if we consider translation between more than one pair of languages. Consider a case where we want to translate among N different languages. The direct approach allows one to exploit knowledge about the particular language pair while developing the systems, but has the disadvantage that we need to create a new system for every new language pair, or $N(N-1)$ such systems. The inter-lingual system is theoretically the best, since it requires only N analysers and N generators. However it requires the creation of a very general inter-lingua, which may be very difficult. The transfer approach lies between these extremes, and is the most widely used approach in practice.

The lexical level deals with looking at the input string of characters and separating them into *tokens*, which may be words, space or punctuation.

Levels of Natural Language Processing

Dealing with natural language typically requires processing at various levels. In increasing order of difficulty, they are:

- ***The lexical level (or the word level)***: This level deals with looking at the input string of characters and separating them into *tokens*, which may be words, space or punctuation. This level also deals with issues like hyphenated words, and misspelt words. It is the lexical level which tells us that the input *"He joined the parti"* consists of four words, of which the last is incorrect. This level is sometimes called *'tokenisation'* or *'lexical analysis'*.

- ***The syntactic level (or the sentence level):*** This level deals with identifying the *structure* of a sentence, and verifying whether a sentence is grammatically correct. This level typically consists of a *'parser'*, which looks at a *grammar* of the language, and the input sentence, and tries to form a *'parse tree'*. If it can form a parse tree, the sentence is syntactically correct, and the parse tree gives us the structure and function of the various components. For example, a typical English sentence would consist of a subject and a predicate. The subject is normally a noun phrase, and the predicate is a verb phrase, and so on. The syntactic level tells us that the sentence *"He the party joined"* is (syntactically) incorrect, even though each word in it is (lexically) correct.

The syntactic level deals with identifying the *structure* of a sentence, and verifying whether a sentence is grammatically correct.

- ***The semantic level (or the meaning level):*** This level deals with the meaning of the input and its components. It is the semantic level which tells us that the sentence *"He ate the*

The semantic level deals with the meaning of the input and its components.

party" is semantically incorrect, though it is lexically and syntactically well-formed. In general, semantic analysis involves knowledge about the world, or at least the relevant aspects of the world.

- *The discourse and pragmatic level (or the conversation context level):* This level deals with information carried across multiple sentences, and with information that is not explicit in the input, but is implicit in the socio-cultural context of the input passage or conversation. For example, the expected answer to the question *"Do you know what the time is?"* is something like *"4 p.m"*, and not just *"Yes"*, though the latter is lexically, syntactically and semantically accurate.

A machine translation system needs to transfer across all these levels.

Issues in Machine Translation

Machine translation (and natural language processing in general), is a difficult problem. There are two main reasons, which are related. The first reason is that natural language is *highly ambiguous*. The ambiguity occurs at all levels – lexical, syntactic, semantic and pragmatic. A given word or sentence can have more than one meaning. For example, the word 'party' could mean a political entity, or a social event, and deciding the suitable one in a particular case is crucial to getting the right analysis, and therefore the right translation. The second reason is that when humans use natural language, they use an enormous amount of common sense, and knowledge about the world, which helps to resolve the ambiguity. For example, in *"He went to the bank, but it was closed for lunch"*, we can infer that '*bank*' refers to a financial institution, and not a river bank, because we know from our knowledge of the world that only the former type of bank can be closed for lunch. To get MT systems to exhibit the same kind of world knowledge in an unrestricted context requires a lot of effort.

When humans use natural language, they use an enormous amount of common sense, and knowledge about the world, which helps to resolve the ambiguity.

Contemporary Machine Translation Systems

The factors mentioned above make it impractical to design completely general-purpose, high-quality, fully-automatic machine translation systems, as already mentioned in the introduction. Contemporary practical machine translation systems therefore adopt one or both of the following strategies:

It is impractical to design completely general-purpose, high-quality, fully-automatic machine translation systems.

1. ***Restrict the domain of application, or the complexity of the language***: This is known as the *sublanguage* approach, which relaxes the criterion of general-purpose translation. One example of a restricted-domain approach is a Canadian system known as TAUM/METEO, which has been translating weather bulletins from English to French, completely automatically, for a number of years. Another example of a sub-language approach is the system in use at the Caterpillar Tractor Company in the US. This company sells tractors in all parts of the world, and needs to maintain manuals in more than 15 languages. They have a system that ensures that the original manuals are written in a subset of English that is sufficiently well-defined to allow automatic translation into all the other languages.

2. ***Involve the human in the loop***: Such systems relax the criterion of full automation, and rely on human editors for pre-editing the input, or post-editing the output, or disambiguating during translation. Depending on who does more work, the systems can be called 'Human Assisted Machine Translation (HAMT)', or 'Machine Assisted Human Translation (MAHT)', also known as *translation tools*. One example of a translation tool is a technical dictionary, which is particularly useful for translation in technical domains, to help ensure correct and consistent translations of technical terms. For example, the word for 'Dialog-box' in Italian is 'finestra', which means 'window', and in French is 'boite', which means 'box'. A translation tool would allow a user who does not know too much French or Italian to translate such terms.

Typical commercial systems (See *Box 2*) use one or more of these methods.

Box 2. Commercial MT System/Tool Examples: Systran and METAL

Systran is an automatic translation system based on the transfer approach. It works on a sentence by sentence basis. The three main module classes are the Dictionary, Systems Software and Linguistic Software. The Dictionary module contains over 2.5 million terms across all covered language pairs, and includes specialised phrase dictionaries and an ability to have user-specific dictionaries. The Linguistic modules contain the analysis, transfer and generation components for the various languages. The Systems module deals with the user interface, file manipulation and integration of the other components.

Systran began as a research project at the California Institute of Technology in 1957, and was later funded by the US government to do Russian to English translation. Over the years, the technology matured and came to be used by corporates like Xerox and Ford. The Altavista Search Engine site uses Systran to offer a translation service. Today, Systran is a set of translation systems ranging from stand-alone PC versions to client-server models, covering 14 language pairs.

For more information, visit the Systran homepage at http://www.systransoft.com/

METAL was a translation system initiated in the late seventies by Siemens-Nixdorf with the University of Texas. It uses the concept of a controlled language to achieve high quality translation in various technical domains. It can also produce indicative translation for general texts, which needs to be post-edited for style. METAL is now called LANT-MARK, and marketed by LANT, a Belgian company. LANT has a suite of related Language Technology products: LANT-Master, a language checker, integrates into existing word processors like MS-Word and allows the vocabulary and style of texts to be in a controlled language which can then be automatically translated; Pangaea is an electronic dictionary that allows the creation of customised phrase lexicons; Eurolang Optimiser, is a translation tool that uses the concept of translation memory.

For more information, visit the LANT homepage at http://www.lant.be/

Current Work

The current focus in MT research is on using machine learning techniques to automatically acquire the lexicon and grammar. This involves using large corpora and applying statistical techniques such as symbolic induction or neural networks to capture correlations in the corpus. The corpora used can be either for a single language, or can be 'aligned corpora' which means a bilingual corpus of translated text in the source and target languages, containing information about which part of the source language text corresponds to which part of the target language text. Another related approach, called as translation memory, is to automatically 'remember' the entire translations

Box 3. Indian Machine Translation Projects

India is a relatively new player in the machine translation field. With its large number of languages, it is a major potential beneficiary of the technology, so research in this area needs to be intensified and coordinated.

Currently, there are at least four major machine translation efforts in India. They are currently being funded by the Technology Development in Indian Languages (TDIL) project of the Department of Electronics (DoE), Government of India.

- The Anusaaraka group with researchers from the Indian Institute of Technology, Kanpur and the Centre for Applied Linguistics and Translation Studies (CALTS) and University of Hyderabad has been working on using a Paninian Grammar approach to deal with Indian languages. The Paninian Grammar approach uses an internal representation that is based on the kaaraka theory developed by Panini to describe Sanskrit grammar, and exploits the similarity among Indian languages.
- The Anglabharati project, also originating from IIT Kanpur, deals with translation from English to Indian languages. It uses a rule-based transfer approach, and has been tested on technical domains, such as product manuals, and medical descriptions.
- The Knowledge Based Computing Systems (KBCS) group at the National Centre for Software Technology (NCST) in Mumbai is currently working on MaTra, a prototype system for human-assisted translation of news sentences from English to Hindi. This group is exploiting the use of human-computer interaction to alleviate some of the traditional problems of machine translation.
- The Centre for Development of Advanced Computing (CDAC), Pune is developing Mantra, a translation system for translating office documents from English to Hindi. It uses a formalism known as XTAG, developed at University of Pennsylvania, and views the translation process as a mapping from one syntactic tree to another.

of frequently occurring phrases or sentences in order to avoid processing them repeatedly.

There is significant activity on machine translation in Japan and Europe, and to a lesser extent, in the US. India is also active in this field, with at least four or five active groups (See *Box 3*). Given the multiplicity of languages in India, such efforts are very relevant, and need to be intensified further.

Conclusion

Two phenomena have given a new impetus to machine translation work – the globalisation of the world economy, and the explosion of the Internet and the World Wide Web. Both these developments mean that there is a need for making an immense collection of natural language documents available to a multilingual global audience, and translation tools and systems can go a long way in meeting that need. The global translation market is estimated to be at least 12 billion dollars. Systems that automatically translate Kalidasa and Shakespeare may still be a distant dream, but systems that translate stock market reports, weather bulletins and technical manuals are a reality today, and will continue to play an increasingly important role in the society of the next millennium.

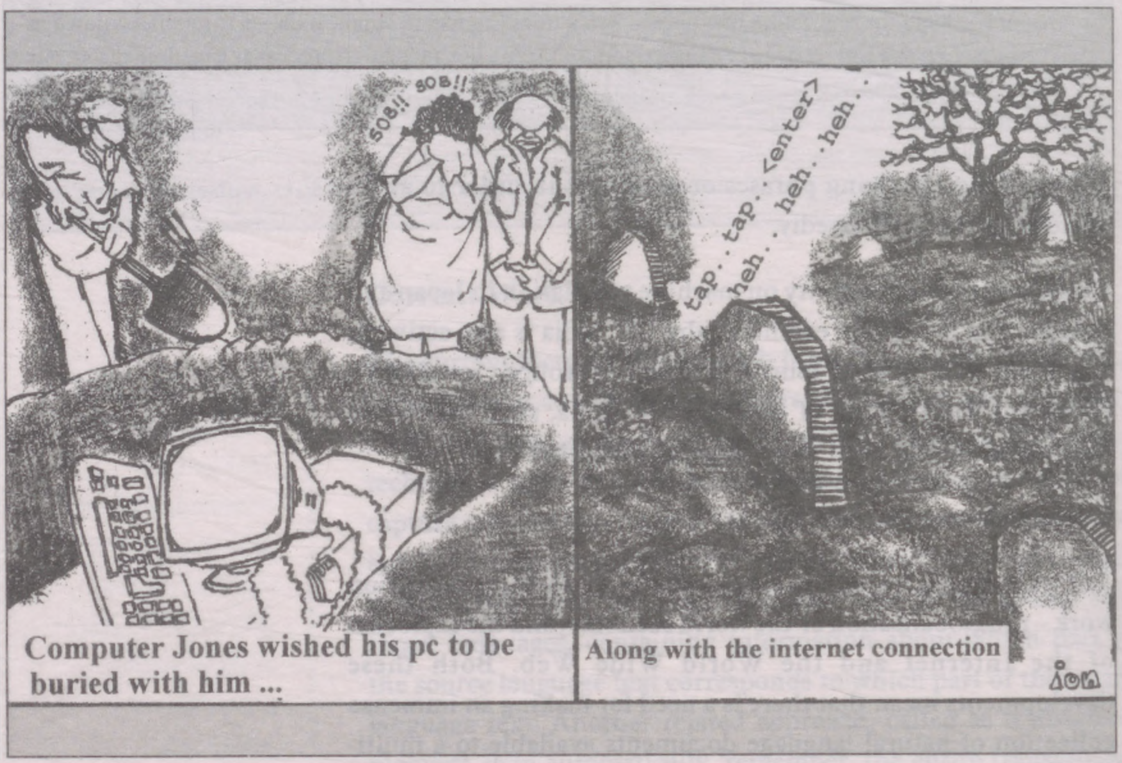
Suggested Reading

[1] The ACL NLP/CL Universe, *http://www.cs.columbia.edu/~radev/u/ db/ acl/html/* Contains a large amount of pointers to on-line NLP and MT information and resources.

[2] John W Hutchins. *Introduction to Machine Translation*. Academic Press, 1992.A good introductory book on machine translation.

[3] The Altavista Service, *http:altavista.digital.com/* This is a good site to search for more information on anything, including Machine Translation. It also runs a simple translation service based on the Systran system.

Address for Correspondence
Durgesh D Rao
National Centre for Software Technology
Gulmohar Road 9, Juhu,
Mumbai 400049
email: durgesh@ncst.ernet.in



Classroom

In this section of* Resonance, *we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.

A Simple and Rapid Method for Isolation of Cellular DNA

N B Ramachandra
Department of Studies in Zoology
University of Mysore
Manasagangotri
Mysore 570 006, India

A simple method has been standardised to isolate cellular DNA rapidly within 1 to 2 hours with common chemicals and simple instruments. This experiment can be carried out at laboratories in undergraduate colleges and high schools.

Introduction

The elucidation of the structure of DNA in 1953 by James Watson and Francis Crick was one of the most exciting discoveries in the history of genetics and molecular biology. After understanding the functional properties of DNA viz, replication, transcription, mutation, recombination and repair, it became possible to manipulate DNA.

In India, we have now introduced a number of new courses at undergraduate and postgraduate levels in life sciences. We teach DNA technology in almost all disciplines of biology. However, the infrastructure to provide practical knowledge of DNA technology to the students is inadequate in most of the colleges due to various reasons. In view of this, I have standardised and demonstrated a few simple and rapid methods to teach recombinant DNA technology at undergraduate and postgraduate colleges by using commonly available chemicals and equipment. Here, I describe the isolation of eukaryotic DNA.

Principle

Most methods of DNA isolation involve the breakage or lysis of the cells to release nuclei and further breakage of nuclei to release the chromatin. DNA in cells exists as nucleoprotein complexes and therefore isolation of DNA involves removal of proteins and carbohydrates (if any) associated with it. Finally, the polymeric nature of DNA is utilised to precipitate it and make it free of small molecular contamination.

Box 1. Merits and Demerits of the Protocol

Compared to standard methods this is a simple and low cost procedure, where a few common chemicals are sufficient for the extraction. The method does not require expensive enzymes like proteinase K and RNAse or expensive equipment. DNA extraction can be completed faster (within 1-2 hours) than by standard methods. This method does not require handling of hazardous chemicals like phenol and chloroform. This method can be introduced at the pre university level and even at the high school levels. Readers are warned that the yield of DNA by this method is low and the DNA extracted by this method cannot be used for research work without further purification.

Reagents, Supplies and Equipment Required

(1) Tissue: spleen/heart/testis/kidney of any vertebrate or coconut endosperm; (2) Mortar and pestles or glass homogeniser; (3) Glass distilled water; (4) Centrifuge (range 3000 to 10,000 rpm); (5) pH meter (optional); (6) 10 ml centrifuge tubes; (7) 30 ml test tubes; (8) Test tube rack ; (9) bent glass rod; (10) Sodium saline citrate solution (SSC–85ml of 0.9% sodium chloride solution + 15ml of 0.5% sodium citrate solution usually gives pH 7.4, if not adjust pH.); (11) 12% Sodium chloride solution (Dissolve 12 gms of sodium chloride in 100ml of distilled water); (11) Absolute alcohol (double distilled alcohol).

Laboratory Protocol

(1) Grind about 200mg of the tissues in about 5ml of SSC in a homogeniser or with a mortar and pestle. (2) Transfer the homogenate into a centrifuge tube and make up the volume to 10ml with SSC. (3) Centrifuge at 3000rpm for 8 minutes and discard the supernatant.(4) Rehomogenise the sediment with 5ml of SSC.(5) Adjust the volume to 10ml, centrifuge at 3000rpm for 8 minutes and discard the supernatant.(6) Then, suspend the sediment in 10ml of 12% sodium chloride solution and centrifuge at 10,000rpm (at least 7000 rpm) for 15 minutes. (7) Transfer the supernatant into a 30ml test tube and add 2-3 volumes of absolute alcohol.(8) Gently mix it by inverting the tube. The white fibrous DNA precipitates. (9) Spool the **fibrous**

white DNA by winding around a clean sterile bent glass rod.

The presence of DNA in solution can be checked by the following methods:

(a) Transfer the spooled fibrous DNA into a 1.5ml eppendorf tube, add 1ml of 70% alcohol, centrifuge for 5 minutes at 10,000 rpm and discard the supernatant. Then the pellet containing the DNA is dried, dissolved in distilled water and optical density is read in a spectrophotometer at 260nm wavelength.

(b) The DNA in solution can be colorimetrically estimated by using diphenylamine colouring reagent [1]. In brief, the deoxyribose purine in DNA in presence of acid forms hydroxylevulinic aldehyde which reacts with diphenylamine to give a blue colour. The formation of blue colour indicates the presence of DNA and intensity of the colour gives the concentration of the DNA in solution.

(c) The DNA is sheared by violent agitation by passing through the small gauged needles as well as by boiling the DNA solution for 10 minutes and chilling it immediately on ice. This is needed to break the high molecular weight DNA, otherwise it cannot get into the gel. For agarose gel electrophoresis, the agarose gel can be prepared by dissolving 0.8% agarose with tris-acetate buffer {4 mM Tris, 2 mM acetic acid, 0.2. mM EDTA, pH 8.1 (TAE)} on boiling and pouring on to the casting tray after cooling the solution to 45°C and placing a slot creating comb before the polymerisation of the gel. Then place the polymerised gel into submarine electrophoretic chamber containing TAE buffer and load the DNA sample into the wells of the gel after mixing with tracking dye. After this connect the power supply and run the gel at 80 volts for 20-30 minutes. Remove the gel and stain with ethidium bromide (et Br), a DNA intercalating dye. The DNA-et Br complex can be seen as an orange coloured fluorescent streaking band under ultraviolet light on a device called transilluminator.

Suggested Reading

[1] R L Rodriguez and R C Tart. *Recombinant DNA Techniques – An Introduction.* Benjamin/Cummings Publishing Company, Inc., 1983

[2] J Sambrook, E F Fritsch and T Maniatis. *Molecular Cloning – A Laboratory Manual.* Second edition, (three volumes). Cold Spring Harbor Laboratory Press. New York, 1989.

[3] B R Glick and J J Pasternak. *Molecular Biotechnology– Principles and Applications of Recombinant DNA.* ASM Press, Washington DC, 1994.

[4] J Jayaraman. *Laboratory Manual in Biochemistry.* Fifth reprint. Wiley Eastern Limited. New Delhi, 1996.

Teaching The Limit Concept

K Viswanath
Department of Mathematics,
University of Hyderabad,
Hyderabad 500 046, India.

"Once more into the breach, dear friends, once more"

Shakespeare, King Henry V, III.1.1

It is the experience of all mathematics teachers that at first contact students face a lot of difficulty with formal analysis. The difficulties are two fold. As Tall says in [1], "First the student usually imagines the definition to *describe* an existing object, rather than *define* the object by deducing its properties, thus finding it strange to 'prove' obvious properties that seem already to be true. Then there are further difficulties because of the *complex use of quantifiers* [emphasis added] and the formality of the deductions. The decision of most UK universities to abandon the teaching of formal analysis as a first year university course is evidence of its huge cognitive difficulty."

Difficult or not, at some stage we have to tackle the problem of teaching formal analysis. The question then is how best we should go about it. At the heart of analysis is the concept of limit, so let us consider how the limit concept could be taught.

To start with we need to decide whether it is possible, and if so whether it is advisable, to present the limit concept in a way which avoids the standard ε–δ formulation. That this is possible has been shown by Hijab in [2]. His approach is the following. Define the limit of a monotone sequence as a supremum or infimum as the case may be. Then define the limit superior and limit inferior of a sequence as limits of appropriate monotone subsequences. When they are equal define the limit of the sequence to be their common value. Finally, define continuity in terms of sequences. The back cover of [2] stresses the fact that ε's and δ's have been done away with and uses it to advertise the book.

There are other interesting innovations in the book and one may want to consider using it as a text for some special batches of students. However in the general Indian context it may not be wise to depart too radically from the traditional approach (say that of Hardy; see [3]). As our students have to seek admission in universities and in the process have to face examiners with varying backgrounds, it may be safer to keep as close as possible to the classical approach.

In [4], Berberian suggests ways of softening the impact of the quantifiers by introducing stepping-stone concepts such as 'ultimately', 'frequently' and 'null sequence'. A sequence is said to be *ultimately* in a set S if all its entries from some point on are in S. It is called *null* if it is ultimately in every set of the form $(-\varepsilon, \varepsilon)$. We then say that a sequence (s_n) converges to s if $(|s_n - s|)$ is a null sequence. A sequence is *frequently* in a set S if infinitely many entries of the sequence are in S. This concept is useful in talking about limit points and in studying the properties of limsup and liminf of a sequence. The approach seems promising, but the student still has to come to grips with quantifiers in order to understand the proofs.

The purpose of this article is to suggest a way of avoiding the use of quantifiers by taking a cue from the subjects of predicate logic and automated reasoning, where use is made of a technical device known as Skolemisation[1] which reduces complex propositions involving quantifiers to a standard form without quantifiers, so that it is in principle easy for a computer to check them. In terms of pedagogy this approach seems a promising one. We briefly present the idea below with some heuristic motivation.

[1] Thoralf A Skolem was a mathematician who contributed to logic, set theory and algebra in the early twentieth century. Suppes in [5] says that what is now known as Zermelo–Fraenkel set theory should in all fairness be called Zermelo–Fraenkel–Skolem set theory.

Let $f : \boldsymbol{R} \to \boldsymbol{R}$ be a function. We think of f as an 'input-output box', with an input of a giving rise to an output of $f(a)$. Now imagine that there is an error in the input a; say we input instead a number x which is close to a. The output would then be $f(x)$ and the error in the output would be $|f(x) - f(a)|$. So an input error of $|x - a|$ has led to an output error of $|f(x) - f(a)|$. It is clearly desirable that we are able to control the error in the output by controlling the error in the input.

Definition. A *control function* for f at a point a is a function $\delta:(0,\infty)\to(0,\infty)$ such that for all $\varepsilon>0$ the following holds:

If $|x-a|<\delta(\varepsilon)$,
then $|f(x)-f(a)|<\varepsilon$.

If f has a control function at a then f is said to be *continuous* at a.

At first glance the definition may seem to be merely a rephrasing of the standard $\varepsilon-\delta$ definition. However some reflection and a little experimentation show that this apparently minor change allows us to present the standard proofs in a logically simpler form. To illustrate this, we present a standard proposition and indicate its proof.

Proposition. Suppose that f and g are continuous at a. Then (1) $f+g$ and (2) $f\cdot g$ are continuous at a.

Proof. Let δ_1 and δ_2 be control functions at a for f and g respectively.

(1) Let $\delta(\varepsilon)=\min\{\delta_1(\varepsilon/2),\delta_2(\varepsilon/2)\}$ for all $\varepsilon>0$ Then δ is a control function for $f+g$ at a.

(2) Let $M_1=|f(a)|+1, M_2=|g(a)|+1$. Define the function δ on $(0,\infty)$ thus:

$$\delta(\varepsilon)=\min\left\{\delta_1\left(\frac{\varepsilon}{2M_2}\right),\delta_2\left(\frac{\varepsilon}{2M_1}\right),\delta_1(1)\right\}.$$

Now suppose that $|x-a|<\delta(\varepsilon)$. Then $|x-a|\le\delta_1(1)$, so $|f(x)|<|f(a)|+1=M_1$. Therefore

$$|f(x)g(x)-f(a)g(a)|\le|f(x)||g(x)-g(a)|+|g(a)||f(x)-f(a)|$$

$$<M_1\cdot\frac{\varepsilon}{2M_1}+M_2\cdot\frac{\varepsilon}{2M_2}=\varepsilon,$$

so the implication '$|x-a|<\delta(\varepsilon)\Rightarrow|f(x)g(x)-f(a)g(a)|<\varepsilon$' holds. Thus δ is a control function for $f\cdot g$ at a.

It is a simple exercise to write the proofs of various propositions

Suggested Reading

[1] Tall D O. Functions and Calculus. in *International Handbook of Mathematical Education Part 1.* Alan J Bishop et al (eds) Kluwer Acad. Pub., Dordrecht, 1996.

[2] Hijab O. *Introduction to Calculus and Classical Analysis.* Springer-Verlag Undergraduate Texts in Mathematics, 1997.

[3] Hardy G H. *A Course of Pure Mathematics.* Cambridge Univ. Press, 1967 (10th edition).

[4] Berberian S K. *A First Course in Real Analysis.* Springer-Verlag, 1994.

[5] Suppes P. *Axiomatic Set Theory.* Van Nostrand, 1960.

[6] Bledsoe W W. Some Automatic Proofs in Analysis. in *Automated Theorem Proving: After 25 Years.* (eds) W W Bledsoe and D W Loveland, Contemporary Mathematics. 29. *Am. Math. Soc.*, 1984.

concerning limits and continuity in terms of control functions. Note that one can similarly define the limit of a sequence in terms of a control function from $(0, \infty)$ to the set of natural numbers $\boldsymbol{N}$.

A case can readily be made in favour of using control functions to teach the limit concept, but whether it actually 'works' (i.e., whether students really benefit) can only be decided after some experimentation by a body of teachers. It is hoped that this article will provide the necessary stimulus for such experimentation.

Please Note

In 'Special Relativity – An Exoteric Narrative: Wherein we put formulas in their place', *Resonance* Classroom Section, Vol.3, No.5, pages 63–72 the figures should appear as follows:

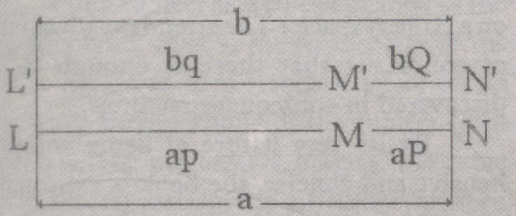


Figure 1.

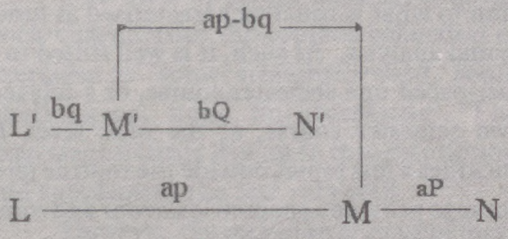


Figure 2.

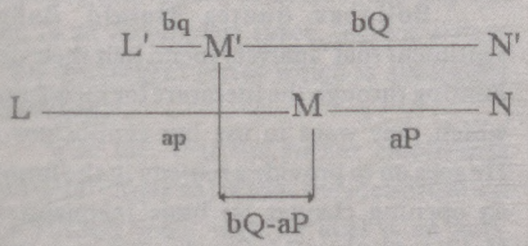


Figure 3.

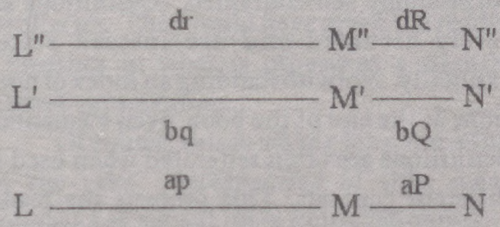
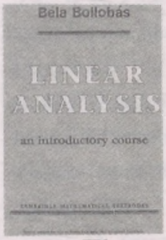


Figure 4.

Linear Analysis

C Varughese



Linear Analysis
Bela Bollobas
Cambridge University Press, 1990
Published in India by Foundation Books
New Delhi, 1994
pp. 240, Rs. 125

The appearance of an Indian edition of Bela Bollobas' book '*Linear Analysis*' must surely be welcomed by advanced undergraduate and masters level students in this country. The author mentions that the book reflects the way he would have liked to have been taught analysis. Many readers will, undoubtedly, echo the same feelings about the work.

The book is a comprehensive introduction to what is usually characterised as functional analysis. As such, it is well suited to a fast-paced one semester course, or a relaxed two semester course, at the MSc level. It includes a few topics outside the routine fare, some to provide a geometric flavour and some to provide a glimpse of more recent results in the subject. The latter serve to give the subjects covered a wide chronological spread.

For a book that is modestly sized, it is surprisingly self contained. Definitions are rarely assumed and are instead stated explicitly. Notwithstanding an index of notations at the end of the book, even elementary definitions are often reiterated when used in different contexts. This protects the reader from the handicap of a definition forgotten.

The author gives proofs for most facts relevant to the main thread of the discussion. These include some, which in similar settings are rarely more than just stated (for example, the well ordering principle is derived from Zorn's lemma). In general, proofs appearing in the book are quite 'efficient'. This is a consequence of a judicious choice and ordering of prior material.

Some of the definitions and results are presented in a more general setting than may be appropriate for a beginner. This is exemplified, for instance, in the rather involved statement of the fact that compact operators form an ideal in the space of bounded operators. Consequently, the onus is occasionally on the reader to adapt the statements to simpler situations to gain a better understanding. (For instance, while discussing an operator between two Banach spaces, one should first visualise the case when both spaces are the same, or possibly an operator on a Hilbert space.) The attendant advantage, of course, is that there is enough to be discovered in subsequent readings.

By its very nature of being comprehensive and concise, the book is expectedly not exhaustive. It is a guided tour of the subject with ample pointers for deeper study. It must be pointed out, though, that no compromise is made on rigour.

Bollobas quotes Harald Bohr's sentiment that 'analysts spend half their time hunting through the literature for inequalities which they want to use but cannot prove'. He goes on to provide a remedy in the form of an opening chapter on basic inequalities... Holder, Minkowski, Cauchy–Schwarz, etc.

The reader is then introduced to normed spaces and linear operators on them.

This leads on to linear functionals and the Hahn-Banach theorem. An interlude on finite dimensional normed spaces is followed by a train of staple theorems...Baire category, closed graph, Tietze–Urysohn, Arzela–Ascoli, Stone–Weierstrass and others. The chapter on contraction mappings that follows might more suitably have been placed later in the book (just before the chapter on fixed point theorems).

A discussion of weak topologies is then followed by a study of Hilbert spaces and some spectral theory. A couple of chapters on compact operators (including the spectral theorem, as a prototype of general spectral theorems) wrap up the operator theory.

A chapter on fixed point theorems and one on the invariant subspace problem (with some known results) provide a finale.

For a subject as well established as the one this book covers, the main body of the text is bound to be a standard set of topics that form the backbone of the subject. But like different performers playing the same musical composition, the difference between authors is the relative emphasis on various results. So it is with this book. Thus we find in it some results that other authors have chosen to play down, or ignore (like the fact that, in an incomplete space, there is always a non-convergent series which is absolutely convergent).

One fact that the author emphasises (and deservingly so) is that the identification of a Hilbert space with its dual is an anti-isomorphism. The belittling of this fact is often a source of confusion. However, the use of the same notation for the adjoint of an operator, acting on the dual space and on the Hilbert space itself (following the identification), could be a mild source of confusion. Stating, for instance, that the spectrum of T^* is the same as the spectrum of T on a Banach space and that the spectrum of T^* is the conjugate of the spectrum of T in the context of a Hilbert space, could be perplexing to a beginner. One hopes that the author's repeated reminders that the identification of a Hilbert space with its dual is antilinear will dispel any confusion.

The notes at the end of each chapter are a pleasant departure from the usual insipid bibliographical listings. These notes provide historical settings for the topics and also the author's polite evaluations (at least the positive ones) of various sources.

The write-up on the back cover of the book characterises the copious collection of exercises as 'some straightforward, some challenging, none uninteresting'. The exercises certainly live up to this description and supplement the text very well.

However, the inclusion of some of these exercises in the main text, as examples, might have served a useful purpose of illustration. It might also have provided a welcome break in the parade of theorems, lemmas and corollaries. This would have given the reader a breather in what is otherwise a slightly intense presentation.

Of course, this is asking for more icing on the cake. Irrespective of this, what Bollobas has served up is quite delightful.

C Varughese, Indian Statistical Institute, Bangalore, India

Reflections

Elementary Particle Physics – Then and Now

The 1951 Indian Science Congress Presidential address (reproduced in the following pages) by Bhabha can broadly be divided into two parts. In the first, Bhabha talks about the development of modern science, the importance of experiments in arriving at the laws of nature and the question of a final fundamental theory. In the second part, he describes the status of 'Elementary Particle Physics' at that time.It may therefore be worthwhile to briefly describe the present situation regarding 'Elementary Particle Physics'.

In early 1951, one already knew about ten different (so called)elementary particles. Soon hundreds more were discovered. This created a crisis since scientists and philosophers have always been fascinated by the idea that the number of basic constituents of nature should be few. Fortunately this crisis was resolved to some extent in the following decades. The modern picture of the basic constituents and their interactions is described by the so called 'Standard Model'. There are four basic forces in nature which in the order of decreasing strength are (i) strong interactions (ii) electromagnetic interactions (iii) weak interactions (iv) gravitational interaction. As for the basic constituents, there are six varieties (flavours) of quarks (u,d,s,c,b,t) each coming in three colours; and six leptons (e, μ, τ, ν_e, ν_μ, ν_τ), plus their anti-particles. Besides, there are twelve gauge bosons, comprising the photon, eight gluons and W^+, W^-, Z^0. According to the standard model, all quarks and leptons are point objects, with leptons (including the electron) experiencing no strong interactions. On the other hand, the quarks and gluons are coloured objects and are permanently confined inside hadrons and cannot exist as free particles. Further, the electromagnetic and the weak interactions are unified into a single force called the electroweak force.This is reminiscent of the unification of electricity and magnetism by Maxwell.

The standard model, though so successful in explaining all the available experimental data, is unable to answer several basic questions including the mechanism for generating the masses for quarks, leptons and massive gauge bosons. Besides, the unification is only partial in this model.In recent years, a 'truly unified' theory called superstring theory has been proposed to unify all the four basic interactions. According to this theory, the basic constituents of nature are not point objects but strings of length 10^{-33} cm. The quarks, leptons and the gauge bosons are merely the different modes of vibration of such strings. The modern unification ideas have brought closer the seemingly contrasting worlds of the very small and the very large. In particular, these ideas hold the promise to explain how the universe evolved after the bigbang, as well as the bigbang singularity itself.

One disturbing aspect of recent theoretical activities is that they are highly speculative, with no experimental data to back them up. It is good to remember Bhabha's views on this issue. As he rightly emphasises, there may be many logically consistent theories which may nevertheless have nothing to do with the actual structure of the physical world.

Are we close to an ultimate theory of everything (TOE) or as others would like to say, a truly unified theory (TUT) ? Again, let us recall what Bhabha says on this matter. According to him, however great the successes of a theory, unless the success is total and complete, it is always possible that something very important may have slipped through the net.

Finally, it is good to remember the words of Bhabha that only science and technology can solve the immense problems facing India. It may be noted here that Bhabha has included both science and technology. This is very important because in this era of 'liberalisation' we think that we only need technology and not science. It would be a grave mistake if we continue to neglect science at the cost of technology. A way must be found so that both can go hand in hand. Only then can this country realise her full potential.

Avinash Khare
Institute of Physics, Bhubaneswar 751 005, India

THIRTY-EIGHTH INDIAN SCIENCE CONGRESS

BANGALORE, 1951

PRESIDENTIAL ADDRESS

CONGRESS PRESIDENT: DR. H. J. BHABHA, F.R.S.

(Delivered on 2nd January, 1951)

The Present Concept of the Physical World.

I express my gratitude to my scientific colleagues from all parts of India for the honour they have done me by electing me to preside over the 38th Session of the Indian Science Congress. A peculiar accident of fate has brought about that the Congress over which I am to preside is being held in this Institute, where I worked for six years from 1940 to 1946 during the period of the last war and I have great pleasure in recording that they were six very happy and fruitful years in my life.

I also wish to express on behalf of all of you, and myself our great appreciation of the fact that our Prime Minister has decided to be present with us on this occasion. That he flew to Bangalore yesterday, and will fly immediately after this meeting to Bombay and thence to England on a mission of prime importance, is a measure of his great personal interest in the development of science in India. Were it not for this, scientific development would receive much less encouragement and support than it does, in spite of the fact that only science and technology can solve the immense problems facing the country, the problems of food shortage, low standard of living and illiteracy.

The multitude and variety of the phenomena of Nature, which still fill us with astonishment, must have bewildered and awed primitive man. It is not strange that he should have sought, on the one hand, to gain some control over them by investing them with anthropomorphic personality which could be influenced by entreaty and prayer, and on the other to alter his immediate physical environment so as to provide some little shelter or margin of safety against the more hostile acts of nature. This urge eventually led to the early civilizations and the later developments following from them. These civilizations depended on a considerable body of practical knowledge acquired empirically, and some highly developed arts and crafts. A few crucial inventions such as that of the horse harness in China, or of the zero in mathematics in India, had a profound influence on their historical development. But with a few notable exceptions, scientific activity in the modern sense did not begin till the Italian renaissance.

Towards the end of the fifteenth century Leonardo da Vinci wrote in one of his manuscripts which is now in the library of the Institut de France (G 96 v),[1]

> "There is no certainty where one can neither apply any of the mathematical sciences nor any of those which are based on the mathematical sciences."

1. The Notebooks of Leonardo da Vinci by Edward MacCurdy, Jonathan, Cape, London 1945.

Reproduced from *Proceedings Indian Science Congress*, Vol.38, 1–11, 1951.

This was not the mere expression of a specialist extolling his own subject, and I quote this sentence because it was written by one who is recognised as perhaps the most versatile genius the world has known, one who had a greater mastery of all the various arts and sciences of his time than anyone since. It expresses the new spirit of the times, a spirit which was to lead eventually to that vast development which is modern science and technology. "The mathematical sciences" for Leonardo consisted of what had been handed down of Greek mathematics, while by the sciences based on the mathematical sciences he understood the applications of geometry to optics and mechanics. What Leonardo wished to emphasize, I feel, was that as long as an observation of a natural phenomenon remained couched in qualitative terms it would not be definite enough to build on, and only by introducing accurate measurement and quantitative relations into it could one be certain that it was right or wrong within the limits of accuracy of the measurements. Some four centuries later Lord Kelvin was to say

> "When you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."

Once this general approach received fairly wide acceptance, the development of science in the modern sense was inevitable.

It was found quite soon that certain properties, which could be stated in terms of exact measurement, were common to many objects. In certain cases, therefore, it became possible to state a general property without specifying the particular object to which it belonged. Such general properties could then be regarded as laws or regularities of nature which all objects of a certain type satisfy. One such regularity or law of nature was the one discovered by Archimedes, that the loss in weight of a body immersed in water is equal to the weight of the water displaced. Archimedes was indeed one of the shining forerunners of modern science. It is nevertheless interesting to note that his law is a law in statics. Laws involving the motions of objects were to come much later. An example of a dynamical law is the regularity discovered by Galileo, that all heavy bodies fall the same distance under gravity in a given interval of time irrespective of their weight. Other regularities of the same type, but which involve more complicated relations between the objects, are the three laws of Kepler on the motion of the planets.

It is important to note that laws or regularities of nature of the type just mentioned are merely empirical statements of properties observed to be common to a large number of objects. They are all unconnected with each other. In order to connect up such regularities with each other it may be necessary to formulate certain more abstract principles or postulates from which the various observed regularities can be deduced.

Newton's fundamental laws of motion exemplify this new approach. Consider his first law, which reads[2]

> "Every body continues in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by a force impressed upon it."

---

2. Cajori, Newton's Principia—a revision of Motte's translation. University of California Press.

Assuming that we understand intuitively what is meant by rest or uniform motion in a right line, we may well ask ourselves what is meant by an impressed force. If we turn to the definitions which Newton has placed a few pages earlier at the beginning of his Principia, we find the answer in Definition IV :

> " An impressed force is an action exerted upon a body, in order to change its state, either of rest, or of uniform motion in a right line."

Expressed in this way Newton's first law would appear to be a tautology : One states that a condition A exists unless interfered with by the existence of B. While B is defined to be present when the condition A is interfered with. It is not the purpose of this discussion to minimize in any way Newton's achievement which is one of the greatest monuments in the history of science, but to understand the real nature of his laws. Let us assume that force can be defined in some other way than in the above definition so as not to make the first law a tautology. One might then imagine the first law to be a statement which we arrive at from direct observation through some process of induction. For example, in commenting on his first law Newton writes

> " Projectiles continue in their motions, so far as they are not retarded by the resistance of the air, or impelled downwards by the force of gravity. A top, whose parts by their cohesion are continually drawn aside from rectilinear motions, does not cease its rotation, otherwise than as it is retarded by the air. The greater bodies of the planets and comets, meeting with less resistance in freer spaces, preserve their motions both progressive and circular for a much longer time."

The inference is that one may conclude by induction that if we could take a body into space to a very great distance from all other material bodies then it would either remain at rest or move in uniform motion in a straight line. We know today that such an induction cannot be made, and may indeed not even be true for the actual world. While it is possible mathematically to assume a world in which Newton's laws are strictly true, it is equally possible mathematically to think of worlds in which they are not. This analysis shows us that strictly speaking *Newton's laws of motion and gravitation are abstract mathematical statements which he quite rightly calls axioms*. And if they came to be regarded as objectively true it is because the behaviour of objects which could be deduced from them by mathematical reasoning agreed with our direct observations. For example, one could deduce from Newton's laws the regularity observed by Galileo concerning the fall of bodies, the three regularities observed by Kepler on the motion of the planets, and a host of other phenomena. Quite appropriately, his epoch-making work was called " The Mathematical Principles of Natural Philosophy."

" The great importance of the contribution of Newton to the development of physics is that it introduced a new approach into science. It led to the acceptance of the position that the ideas which are to be regarded as fundamental for the understanding of nature are certain abstract concepts or postulates which cannot be proved directly, and not the directly observable regularities of nature which can be deduced from them. This position was accepted because it allows one to order different empirically found regularities of nature into a unified logical scheme which would not otherwise be possible."[3]

3. H. J. Bhabha, Presidential Address to the Section of Physics, Indian Science Congress, Calcutta, 1943.

" A consequence of this approach is that any newly discovered fact of nature which does not fit into the existing scheme of physics may necessitate a complete change of the fundamental postulates. Since, however, the old postulates were such that a very large body of observed facts about nature could be deduced from them, it follows that they must still have a restricted validity under certain circumstances, and be deducible as approximations from the new postulates. Although, therefore, every new discovery which does not fit into the old scheme necessitates a complete change of the fundamental postulates, the change is always from a certain set of concepts to a set of more general concepts. As one goes deeper and deeper into the understanding of nature by co-ordinating all the known facts into one scheme by the use of wider concepts as the basic postulates, the old fundamental postulates become, in a sense, a part of the superstructure, taking a place in between the new fundamental concepts and the directly observed regularities of nature.

As an example of this process of generalization of the basic concepts, one may recapitulate the well-known development from the pre-relativity concepts of an absolute space and an absolute time to the more general concept of the unified space-time of the theory of relativity. In pre-relativity physics, in recognition of the arbitrariness of the orientation of the three axes of the frame of reference, the natural laws were formulated so as to be invariant for all rotations of the space axes. Time, on the other hand, was assumed to be absolute and the same for all observers. However, in consequence of the observation that the velocity of light c is the same for all observers in uniform motion relative to each other, the idea of absolute rest has had to be discarded leading to the principle of relativity, which demands that the laws of nature should be so formulated as to have the same form for all observers moving relative to each other with uniform velocity. Stated mathematically, the special theory requires that the fundamental equations shall be invariant for all transformations of the Lorentz group, whereas in pre-relativity physics the laws were only invariant for all transformations of the three dimensional rotation group, which is a subgroup of the Lorentz group."

The above example also serves to show how the basic concepts of a theory may be radically changed, while still retaining most of the notions of the earlier theory, but recognising them to be of limited validity, true not universally but only in certain circumstances. Thus, the absolute distinction between a time interval and a space interval in pre-relativity physics is replaced in relativity theory by the absolute distinction between time-like and space-like intervals, while the notion of the absoluteness of time of the earlier theory is seen to be approximately correct in the new theory for a group of observers moving relative to each other with velocities small compared with that of light.

A widening of the basic concepts automatically reduces the amount of arbitrariness in the theory. For example, in pre-relativity theory the force between two bodies could be taken to be entirely arbitrary. In relativity theory, on the contrary, the force has necessarily to be conveyed through the medium of a field. The basic differential equations which any field has to satisfy in relativistic theory are drastically restricted in their variety by the same requirement of relativistic invariance, so that there is a very limited freedom in the choice of the form of the force which can be exerted between two particles.

When a science reaches an advanced stage, as physics undoubtedly has today, the facts which can be discovered by direct observation become more and more meagre. We may expect, for example, to be able to discover by experiment the masses of the various types of elementary particles, the different processes which they undergo, their general behaviour in passing through substances of different types and so on. It is, however, inconceivable that an equation like the Dirac equation could be deduced by direct observation in some such way as Maxwell deduced the equations of the electromagnetic field. There is, therefore, no other path open to us but to proceed along the lines I have indicated above. In such an advanced stage however we have certain compensating advantages. We have a number of theories to fall back upon with the knowledge that each correctly describes a large body of experimental evidence in certain circumstances. We can, therefore, attempt to proceed by evolving new theories which reduce to the previously known ones in the circumstances in which the latter are known to be correct. As I have stated on a previous occasion then " The aim of theoretical physics must be to find a complete set of mutually consistent mathematical postulates or axioms from which the properties of nature, meaning thereby the result of every conceivable experiment, can be deduced in the form of a series of theorems. It is, however, necessary in order to achieve the last step of comparing the mathematical statements of the theorems with the results of observation that the basic mathematical postulates must be supplemented by a set of prescriptions about the interpretation of the mathematical formalism. It is clearly not sufficient that the postulates should be consistent and their correctness from the point of view of physics can only be demonstrated by an agreement between the deductions and the results of experiment."[4]

It is most important to distinguish this approach from the one which assumes that one can arrive at the laws of nature by pure thought and epistemological reasoning. The latter approach has neither met with much success, nor proved particularly fruitful in promoting an understanding of the physical world. In our approach, on the contrary, we recognized that it may be possible to build many logically consistent theories which have nevertheless nothing to do with the actual structure of the physical world. Theories in pure mathematics provide many such examples. If any set of axioms or postulates can claim to correspond to reality it is because the deductions from them stand the test of agreeing with the results of experiment.

We must turn now to review the development of our picture of the physical world resulting from recent discoveries. It had already been established by the end of the last century that the multitude of substances in nature are all made up by the chemical combinations of a certain number of basic substances called the chemical elements. The smallest unit of a given chemical element was called an atom. The combinations of these atoms, either of the same element or of different elements gives rise to chemical compounds, which compose the body of all the substances that we meet in nature.

Investigation on the conduction of electricity through gases led Thomson towards the end of the last century to the discovery of the fact that this conduction could be attributed to a particle of negative charge having always the same ratio of charge to mass irrespective of the substance

4. H J. Bhabha, Reviews of Modern Physics, 21; 451, 1949

under investigation. Moreover, this mass was some thousand times smaller than that of a positive ion. Subsequent researches have established the fact that there is a smallest unit of negative electricity, smaller sub-divisions of it not being found in nature, and that all negative electricity appears in integral multiples of this smallest unit, which is now denoted by e. Thus it came to be established that there was a type of particle, called an electron, which always possessed the same negative charge e and the same mass m, which was somehow contained in atoms, and whose behaviour was responsible for the phenomenon of electricity.

Since an atom is an electrically neutral body, it follows that if electrons are contained in it, then it must also contain an equal amount of positive charge. It was not clear at the time how the electrons and the positive charges of electricity were distributed in an atom. For example, were the electrons embedded in a uniform medium of positive electricity, rather like plums in a cake ? Or were they like planets revolving round a sun of positive charge ? The answer to this important question was furnished by Rutherford in 1911. He showed by a study of the scattering of α particles that the true picture of the atom was to consider it like a solar system in which the electrons move like planets round a heavy centre called the nucleus in which all the positive charge and most of the mass of the atom is concentrated. Since the negative charge inside the atom depended on the number of electrons in it and was an integral multiple of e, the positive charge on the nucleus had likewise to be an integral multiple of e. It was soon established that the number of units of positive charge on the nucleus determined the chemical properties of the atom, and that there were 92 such chemical elements ranging from the lightest, hydrogen, to the heaviest, uranium, with 1 to 92 positive units of charge on the nucleus respectively.

The mass of the nucleus of the lightest element, hydrogen, containing just one unit of positive electricity, was found to be always precisely the same, and some 1840 times the mass of the electron. Since this nucleus of hydrogen never broke up into smaller fragments, it became convenient to regard it as a new type of fundamental entity, a new elementary particle, called a proton.

Further researches showed that the mass of any atom was always almost precisely an integral multiple of the mass of the proton, while its charge was a smaller integral multiple of the charge of the proton. These facts led one at the time to accept a picture of the nucleus which made it appear to be made up of protons and electrons only. The number of protons was sufficient to make up the mass of the nucleus, while a certain number of electrons were added inside the nucleus to neutralize the charge of some of the protons and make the total positive charge equal to the actual charge of the particular nucleus. Thus round about 1930, our picture of the physical world appeared to be remarkably simple. The whole material world was thought of as made up of just two types of elementary particles, protons and electrons. By suitable arrangements of these one built up the atoms of the chemical elements. And from suitable arrangements of the latter every other material thing that was found in nature. Light, or in more general terms, electro-magnetic radiation, or photons, and gravitation, were the only two other physical entities found in nature.

A scientist at that time could have thought, as many did think, that when one knew the mathematical laws governing the behaviour of these four elementary types of physical entities, the protons, electrons, photons and gravitation, one would know everything of a fundamental nature that there was to know of the physical world, and physics in principle would be a subject which had reached its destination. The subsequent development of the last twenty years shows us how far this belief was from the truth. It shows in a striking manner that however great the successes of a theory, unless this success is complete and total, it is always possible that something very important may have slipped through the net. The apparently small but persistent difficulties or inconsistencies in a theory, or small discrepancies between theory and observation, may be essentially unbridgeable within the framework of the basic concepts of that theory and yield the clue to new ideas.

What were these difficulties of which I have spoken ? In order to make a body spin about itself like a top we have to impart to it energy and something called angular momentum. It is found that like electricity, which occurs in nature only in integral multiples of the basic unit e, so angular momentum also occurs only in integral multiples of a basic unit which is just Planck's constant divided by 4π, that is $h/4\pi$. Spectroscopic analysis has shown that the two elementary particles, the proton and the electron, each possesses an intrinsic angular momentum or spin of one unit which arises, so to speak, from its spinning about itself like a top. On the other hand, angular momentum which arises from one body moving bodily round another is always an even integral multiple of the basic unit, that is either zero or two or four etc. times $h/4\pi$. Spectroscopic analysis shows that the spin of a nucleus containing an odd number of heavy particles, that is an odd number of protons in our picture, is always an odd multiple of the basic unit much as if the electrons in the nucleus did not contribute to the spin at all, unlike the electrons outside the nucleus each of which must contribute an odd number of units due to its intrinsic spin and its bodily motion. Secondly, both protons and electrons satisfy a law which the theoretical physicist calls Fermi Dirac statistics. It can be shown then that a nucleus containing an odd number of protons plus electrons must also satisfy the same statistics so that in a molecule composed of two such atoms only certain spectral lines must appear and not others. Experiment again shows that the statistics of such nuclei appear to depend only on the number of heavy particles in the nucleus and not on the total number of protons plus electrons in our picture. In fact all the nuclei seem to behave as if the electrons which were supposed to be in them only manifested their electric charge but neither their spin nor their statistics. A bold attempt to face this difficulty would soon have led one to the view that nuclei were not composed of protons and electrons but rather of protons and some hitherto unknown particle having to a very high degree the same mass as the proton, the same spin and satisfying the same statistics. A particle of this description was discovered by Chadwick in 1931 and was called a neutron. It had to be accepted as a new elementary particle and not a composite structure made up of a proton and an electron for the same reason that prevented us from thinking consistently of the nucleus as being made up of protons and electrons. It immediately led to the acceptance of the picture that all nuclei are composed of only two types of particles, protons and neutron. The number of types of elementary particles was thus increased by one.

The acceptance of the neutron as an elementary particle, however, introduced a new feature into our concept of the elementary particles. For it had been known for a long time that certain nuclei, as for example those of the radio active elements, emit electrons every now and then. One can only fit this fact into the picture by assuming that when such an electron is emitted from the nucleus it is in fact newly created in the process and that simultaneously a neutron in the nucleus changes into a proton. Thus we have to admit the possibility that while the elementary particles are not composite, and that as long as they exist they are immutable with absolutely constant properties, nevertheless there are occasions when one or more such particles can disappear altogether with the simultaneous creation of another set. For example, a neutron may disappear and give place to a proton and an electron. Since the neutron, proton and electron all have a spin $h/4\pi$ and the bodily motion of these particles can only contribute an even multiple of $h/4\pi$ the conservation of angular momentum and statistical properties compels us to postulate that there must be yet another elementary particle called, a neutrino by Pauli, which possesses no charge, a mass negligible compared with that of the electron and a spin of one unit ($h/4\pi$).

In 1931, Anderson reported a photograph which seemed to be that of a particle of the same mass as an electron and having one unit of positive instead of negative electric charge. The experimental advance of making the cosmic rays themselves take their own photographs instead of taking photographs at random in a Wilson Chamber then enabled Blackett and Occhialini soon afterwards to discover a new phenomenon called cosmic ray showers. Although cosmic rays are a relatively rare event Blackett and Occhialini showed that very frequently many such rays occurred in a shower and subsequent work has demonstrated that such showers of particles are produced by cosmic rays when they pass through matter, as for example sheets of lead placed in the Wilson Chamber. Blackett and Occhialini showed that their showers contained not only the usual electrons but a comparable number of electrons with the opposite charge. With this, the existence of the positron, as this new particle was called, was established.

The existence of the positron could be understood immediately in terms of an equation for the electron which Dirac had put forward in 1928 and which combined in it for the first time the ideas underlying the theories of relativity and quantum mechanics. Dirac had already shown that certain apparent difficulties in his theory could be understood as expressing on the one hand the existence of a particle of equal but opposite charge to that of the electron and on the other the possibility of a pair of such positive and negative particles being created by the materialization of energy or of their annihilation with the transformation of their mass energy into radiation. Subsequent experiments have fully confirmed the correctness of these basic processes predicted by the theory. Nevertheless, a consequence of this theory was that no electron or photon of even the highest energy could penetrate large amounts of matter, while a growing body of evidence from cosmic ray experiments indicated that particles which looked like electrons did in fact penetrate great thicknesses of matter. Thus, there seemed to be evidence that quantum theory failed for very high energy electrons, while at the same time there was no theory to explain the phenomenon of the cosmic ray showers. It was only when the Cascade Theory put forward by Heitler and the present author showed that the existence of cosmic ray showers and the behaviour of the soft component of cosmic rays in the atmosphere and in dense substances could be explained on the basis of quantum theory was it possible to conclude that the electron-like tracks of particles which did not behave completely like electrons nor like protons must be due to a new type of particle having an intermediate mass. Thus, the existence of a new particle called the meson, with a mass some 204 times that of the electron came to be established in 1938.

Please Pass on This Form to an Interested Friend/Colleague or Use This to Order a Gift Subscription to Someone in India or Abroad❖

Resonance

journal of science education

ORDER FORM

YES! send issues of *Resonance* at the address given below.

Please tick the correct boxes

INDIA			FOREIGN (12 issues)	
☐ Institutional †	☐ 12 issues	(Rs. 400)	☐ Third world countries	☐ Institutional (US$ 50)
	☐ 24 "	(Rs. 700)		☐ Personal (US$ 25)
	☐ 36 "	(Rs. 1000)		
☐ Personal	☐ 12 issues	(Rs. 200)	☐ Other countries	☐ Institutional (US$ 100)
	☐ 24 "	(Rs. 350)		☐ Personal (US$ 50)

For renewal indicate the subscription no. ....

☐ DD (No dated)

☐ Money order ....

Name ....

.... (Signature)

Address ....

....

....

.... (Date)

Pin ....

Please note:

- ❖ We accept rupee payment equivalent to US dollars and send the journal abroad by air mail.
- Drafts/Cheques are to be drawn in favour of INDIAN ACADEMY OF SCIENCES
- For Indian subscription, Demand draft or Money order is preferred. We also accept cheques in case of genuine difficulty. **For outstation cheques, please add Rs. 20/- towards clearance charges**.
- Foreign subscription rates include air mail postage
- For renewals, please quote subscription number

Use xerox copies if more order forms are required

† Accepted for calendar year (January–December) only

Address for Correspondence: *The Circulation Department, Indian Academy of Sciences, C V Raman Avenue, P. B. No. 8005, Bangalore 560 080, India.*

The Circulation Department
Indian Academy of Sciences
P. B. No. 8005, C V Raman Avenue
Bangalore 560 080

A particle with a mass of this order of magnitude had already been envisaged by Yukawa in 1935 in an attempt to explain the short range nature of nuclear forces, that is the forces between two particles in the nucleus of an atom, as for example a proton and a neutron. The observed mesons, therefore, came to be regarded as the agency responsible for nuclear forces, and in accordance with this picture the beta decay was then considered as due to the decay of virtual mesons emitted by nuclei. Research carried out since the end of the war has demonstrated that the picture was again not as simple as it was then supposed to be. Firstly, although the decay of mesons into electrons has been confirmed by experiment, more accurate experiments have shown that the electrons are emitted with a continuous distribution of energies and not with a sharply defined energy as originally supposed. This inevitably leads to the conclusion that in the process of the decay of a meson into an electron not one but two neutral particles must be emitted.

Secondly the identification of the observed cosmic ray mesons with the particles responsible for nuclear forces inevitably requires them to have a strong interaction with nuclei, whereas the mesons observed in cosmic rays were seen to penetrate large amount of matter with but a very weak interaction with nuclei. A serious attempt to face this difficulty might easily have led one to the conclusion that the particles responsible for nuclear forces were not in fact the observed mesons. This conclusion was, however, not accepted until the studies by Powell and his group of the tracks of cosmic ray particles in special photographic plates had shown that there are in fact two types of mesons with two different masses, and that the one type decays into the other with a period of about one hundred millionth part of a second. The two types of mesons are now known as pi and mu mesons respectively. The meson generally seen in cloud chamber photographs are the mu mesons whereas the pi mesons are the ones which are now identified as being responsible for nuclear forces. In accordance with this picture one would have to attribute an even integral spin (in units of $h/4\pi$) to pi-mesons. The mass of pi-mesons, or pions as they are called in short, as determined from a study of the density of the tracks they produce in special photographic plates appears to be in the neighbourhood of 286 times that of the electron.

More recent experiments with the large cyclotron at Berkeley and elsewhere have led to the discovery of yet a new elementary particle a neutral pion, that is a pion with practically the same mass as the charged pion but with no electric charge. Although such a neutral pion interacts strongly with nuclei, nevertheless it cannot be observed directly in a photographic plate or in a cloud chamber due to the fact that it does not possess an electric charge and therefore does not ionize. Despite this, however, it has been possible to ascertain its mass with very great accuracy. The reason is that a neutral pion decays spontaneously in a time of the order of a hundred million millionth part of a second into two gamma rays. Experiments at Berkeley have shown that when a negative pion hits a proton the latter is transformed into a neutron with the emission of a quantum of radiation carrying away the entire mass energy of the pion, that is some 140 million electron volts. But there is another alternative which can result from this collision, namely the conversion of the proton into neutron with the emission of a neutral pion. The neutral pion then decays immediately into two photons with roughly half the energy of the original pion namely seventy million electron volts. However, the neutral pion that is emitted decays while in motion, thus resulting in a certain spread in the energy of the two gamma rays. From this spread one can calculate the kinetic energy of the emitted neutral pion and from this again the difference in mass between the neutral and the negative pion. In this way one finds that the mass of a neutral pion is only a few electron masses less than that of a charged pion.

Rossi has shown that in cosmic rays some two-third of the total energy is converted into charged penetrating particles while one-third disappears into neutral charged particles. Since charged pions can have a positive or a negative charge, it follows from this that the interactions of a positive, a negative or a neutral pion with a nucleon are roughly of the same magnitude.

I now come to particles whose existence is highly probably though not absolutely certain. In 1947 Rochester and Butler in Blackett's laboratory reported certain unusual events which they had observed in a Wilson Chamber. They occasionally saw the tracks of charged particles which seemed to show an abrupt change in their direction in the gas of the chamber. They showed that it was difficult to interpret these tracks as due to scattering in the gas for two reasons. Firstly no recoil nucleus in the gas was visible. Secondly it would be difficult to understand why a particle should have such a great likelihood of collision while passing through the rarified medium of a gas and yet pass through dense matter like a lead plate without suffering any collision at all. They put forward the explanation that these forks were due to the spontaneous decay of a charged particle into another charged particle and a neutral one. Experiments at present being carried out at Pic-du-Midi and Jungfraujoch are rapidly producing further evidence that this interpretation is correct. It would be consistent with the present evidence to interpret the charged particle resulting from the decay as a pion. Whether all the original particles are of the same type is not a question that can be answered at present. Most of the particles appear to have a mass in the neighbourhood of 800 times the electron mass. But there is some indication that there may also be particles of this type with masses more than a thousand times that of the electron. Rochester and Butler and more recently Butler and his collaborators have also produced evidence to show that neutral particles of corresponding mass exist which seem to decay into two charged particles in the same way.

Lastly one should mention the case of a particle observed by Powell and his group and called by him a tau meson, which came to the end of its range and emitted three mesons of which one is certainly a pion.

We see now that at least nine different types of elementary physical entities exist in nature, while the existence of two more is almost certain. While experiments may give us information about the masses of these particles, their mutual interactions and the processes in which they take part, it seems inconceivable that an experiment would enable us to deduce directly the mathematical equation describing the behaviour of any such particle. We can only hope to set up the mathematical equations governing the behaviour of these particles by taking as our guides certain well known principles, as for example the principle of relativity and the ideas underlying quantum mechanics. Even such a clearly defined property as the spin of an elementary particle is not something we can hope to measure directly in the case of particles like the meson but must infer it from considerations of the processes in which they take part, by comparing the behaviour of particles of different spins as predicted by theory with the experimental observations.

The circumstance that there are a dozen different types of elementary particles in nature would lead us to expect that there may be many more, and indeed with our present knowledge we cannot exclude the possibility that there may be an infinite number of them. This does not mean, however, that we shall never be able to obtain a complete description of them all. There are, for example, an infinity of lines in the spectrum of hydrogen and yet we possess today not only a formula which in one neat expression contains the energies of all these lines but also a mathematical theory which allows us to calculate all the stable states of the hydrogen atom, other properties such as the scattering of electrons by atoms, their creation by photons, and even more complicated properties like the nature of the chemical bond between two hydrogen atoms. It is, therefore, quite possible that with increasing knowledge we may be able to find the formula which gives us the masses of all the elementary particles and the general principles which will allow us to deduce the equation satisfied by a particle of any particular mass.

Lorentz at the beginning of this century regarded the charge of an electron as a property of the electron and tried to explain its mass as due to the energy of the electromagnetic field associated with that charge. The idea that the mass of the elementary particle is wholly of field origin has had to be abandoned today because we know a number of elementary particles all having the same electrical charge but different masses. On the other hand one is faced with the fact that whenever the electromagnetic field interacts with any other type of physical entity, be it an electron, a meson or a proton, then the measure of this interaction, namely the charge of the particle, is always the same. From a phenomenological point of view, therefore, we would be more justified today in considering the electric charge e of an elementary particle as a property of the electromagnetic field rather than of the particle, while considering the mass of the particle as an intrinsic property of the particle, unconnected with its interaction with the electromagnetic field. Our approach to this problem today should therefore be just the opposite of that of Lorentz. If the electric charge e is to be considered as a property of the electromagnetic field, as I have suggested, then since the only unit associated with the field in which it could be measured is the square root of Planck's constant multiplied by the velocity of light, we should consider this ratio, or its square e^2/hc to be an intrinsic property of the electromagnetic field. The dimensionless constant e^2/hc would then appear to be a number associated with the electromagnetic field and not a universal constant of nature, of the same status as Planck's constant h or the velocity of light c which enter into the description of other elementary particles.

It is clear that we are now penetrating into a new level of nature which was practically unknown some twenty years ago. I have pointed out earlier that although there may be an infinite number of types of elementary particles nevertheless this fact in itself does not necessarily force the conclusion that we will never be able to describe nature fully or to explain the physical world exhaustively. On the other hand we cannot be certain with our present knowledge that a complete mathematical theory of the physical world can be based upon a finite number of postulates, and if this were not so we would be faced with a situation in which we could never hope to give an exhaustive description of everything there is in nature, but only to extend with the flow of time the region which we had explored and understood.

Information and Announcements

Glimpses of R & D Work in BARC

It was in December 1939 that the existence of nuclear energy came to be known to the world when the first paper on nuclear fission reaction was published in *Nature*. Even before the fact that the feasibility of a sustained fission chain reaction was demonstrated in Chicago was made known to the world, Homi Jehangir Bhabha initiated efforts in March 1944 to start nuclear research in India so that when nuclear energy is successfully applied for power production, India will not have to look abroad for experts but will find them ready at hand. The Tata Institute of Fundamental Research was started in 1945 with Bhabha as its Director. After India gained independence, on March 23, 1948 Prime Minister Pandit Jawaharlal Nehru introduced in Parliament a bill to develop and promote the use of atomic energy. While introducing the bill Pandit Nehru remarked "because of the powerful tools that atomic energy provides for unravelling and understanding the processes of life, because of the new weapons it gives for fighting disease and for the alleviation of human suffering, and because of the concentrated source of power it puts in our hands for peaceful purposes, man took one of the greatest strides forward in history by discovering how to release atomic energy." This bill was passed and became the Atomic Energy Act of April 1948. The Atomic Energy Commission was set up in August 1948. In January 1954 the Atomic Energy Commission decided to set up the Atomic Energy Establishment, Trombay which was

renamed as Bhabha Atomic Research Centre in January 1967.

The first scientist whom Bhabha recruited for the Atomic Energy Commission was a biologist. This illustrates the importance of biological research in the peaceful applications of nuclear energy. One of the first facilities that was set up at Trombay was the 1MW research reactor, Apsara, which became operational in 1956. Research in different branches of physics, chemistry and metallurgy were initiated as a part of the atomic energy programme under the aegis of TIFR even before the Atomic Energy Establishment, Trombay came into formal existence. Electronics is another area which is essential for the application of nuclear energy for any kind of nuclear research.

After Apsara, which is still in operation, BARC went on to build a zero energy experimental reactor named ZERLINA, a 40MW research reactor, Cirus, a plutonium fuelled fast reactor, Purnima-I, a 100 MW high flux research reactor, Dhruva, a Uranium–233 solution fuelled critical facility, Purnima-II, and a Plutonium fuelled mock up facility, Purnima-III, at Trombay, which was installed as a 30 kw neutron source reactor called KAMINI at the Indira Gandhi Centre for Atomic Research at Kalpakkam. The ZERLINA reactor was primarily meant to be a facility where the neutronics behaviour of different types of reactor lattices would be studied in order to refine the analytical tools used in the design of large research reactors and power reactors. After successfully using ZERLINA reactor for this purpose for about 20 years it was de-commissioned in 1984.

Basic studies aimed at gaining better understanding of nuclear fission and other nuclear reactions, use of neutron diffraction and neutron scattering to study a multitude of characteristics of different materials, irradiation of nuclear fuel assemblies in order to evaluate the performance of the fuel assemblies in a power reactor and irradiation of various materials in order to produce a variety of radioactive isotopes are some of the purposes for which the research reactors at Trombay have been used.

Physical Sciences: In BARC, there are strong research groups in a number of areas such as nuclear physics, lasers, solid state physics, crystallography, reactor physics, spectroscopy, seismology and gamma ray astronomy. The research reactors and the accelerator facilities built and operated by this centre have provided the foundation for the basic research that is carried out in many frontier areas of nuclear physics and condensed matter physics. BARC is also in the forefront of designing and building complete state of the art computer instrumentation for its various programmes in basic and applied research in physics.

Experimental nuclear physics programme utilises the Dhruva and the Cirus reactors, the 5.5 million volt van de Graaff accelerator, the 2 MV Tandem accelerator, the variable energy cyclotron in Calcutta and 14 MV pelletron accelerator at TIFR. Theoretical investigations to elucidate the structure of nuclei and for understanding a variety of reaction mechanisms are being carried out.

The original van de Graaff accelerator at Trombay was converted recently into a 7 million volt folded tandem ion accelerator (FOTIA), for more advanced studies. In condensed matter research, investigations using neutron beams from the Dhruva and the Cirus reactors constitute the major programme. A number of techniques have been developed. These include X-ray diffraction, laser Raman scattering, Mossbauer spectroscopy, gamma ray Compton scattering, positron annihilation, and Auger electron spectroscopy.

Using neutrons and X-ray beams, structural investigations of a variety of solids are being carried out, including those of biological importance. Neutron beams are also being used for the study of phonon spectra, elucidation of magnetic structures of solids and for dynamical studies of liquids and molecular systems. Positron annihilation and Compton profile studies are being carried out to understand electronic states in solids. Mossbauer spectroscopy is being employed for the study of cooperative phenomena as well as for corrosion studies. Laser Raman scattering investigations are being conducted for elucidating the nature of phase transitions in many solids. These studies are being supplemented by other techniques such as birefringence measurements and neutron scattering.

Facilities have been developed in BARC for carrying out investigations on solids under high pressure and on accompanying phase transitions. Theoretical studies have provided a new model for the equation of state in the difficult intermediate pressure region of 5 to 100 Megabars. The Centre also possesses good cryogenic facilities such as helium liquifiers, superconducting solenoids and cryostats to extend some of the measurement techniques to very low temperatures.

A variety of new devices and spectrometers for utilising the neutron beams available from the Dhruva reactor were built. These include hot and cold neutron sources, a guide tube laboratory and automated neutron spectrometers. A hallmark of condensed matter research activity has been an intense effort to indigenously develop sophisticated complex experimental equipment including lasers, computer controlled diffractometers and neutron spectrometers.

Excellent facilities are available at Trombay for high resolution spectroscopic studies. These are being utilised for the study of a variety of molecules in addition to spectral studies of rare earth ions. New band systems have been obtained for many diatomic and simple polyatomic molecules in microwave discharge and by flash photolysis. Fine structure analysis of the band systems has led to accurate evaluation of rotational constants. A vacuum ultraviolet Saya–Namioka monochromator has been built and is being used for carrying out spectroscopic studies of excited ions using tandem accelerator.

Chemical Sciences: Basic research in the field of chemical sciences is directed towards areas such as chemical dynamics, radiation and photochemistry, laser chemistry, interfacial phenomena, catalysis, radiation damage studies on organometallics, high pressure studies on materials, transport properties of metal oxides and hydrogen storage materials.

The interaction of high energy radiations like gamma rays, electrons and alpha particles in the MeV energy range with materials in aqueous and organic environment is a frontier area of both fundamental and applied importance. Work in this area has resulted in the development of new routes for radiation synthesis, catalysis, polymerisation, vulcanisation of natural rubber, etc. Development of radiation gelled and encapsulated biomedical devices for diagnosis and blood purification, pressure sensitive adhesives and diamond film deposition techniques are some of the significant achievements.

Dynamics of ultrafast reactions ranging from billionth of a second to trillionth of a second are studied to determine how energy deposited in a chemical environment is utilised for the purpose of chemical reactions, how the reactions themselves evolve and go through several intermediate stages and in what way the reactions and their energetics can be channelised in specific directions.

Development of high purity materials forms an integral part of a successful nuclear energy programme.

Analytical methodology has been developed for estimation of impurities present in parts per million or lower concentration in materials of interest in the nuclear programme using advanced techniques like neutron activation, mass spectrometry, atomic absorption and electroanalytical methods. Application of nuclear analytical methods to geochemistry, forensic science and for ultratrace analysis forms an important part of this work.

The continuing emphasis on both fundamental and applied aspects of chemistry has resulted in important spin-offs such as instruments based on gas chromatography including automatic process analysers, microsecond flash photolysis apparatus, temperature jump apparatus for chemical relaxation studies, electrochemical instruments and nuclear instruments for nondestructive assay, processes pertaining to heterogeneous noble metal catalysts for heavy water production, radiation processes for production of superfine metal powder, novel polymers and polymer composites including radiation grafting, and use of laser photochemistry for isotopic enrichment.

Basic studies in all branches of science and engineering are encouraged in BARC. Dynamics of ultra fast reactions taking place in billionth of a second, gas chromatography, development of bio-pesticides and food preservatives are all areas in which scientists and engineers work in BARC. Development of superconductor based magnets, liquid crystals, linear accelerators for medical and other applications, low temperature physics, memory alloys, biotechnology, development of new radiopharmaceuticals etc. are some of the fields in which BARC scientists work.

Computers are essential for any scientific work in the present day world. Getting numerical and analytical solutions to problems, simulation of various phenomena, and data capturing and processing need fast computers. Starting from indigenously available microprocessors, and using ingenious architecture, computer scientists in BARC developed a parallel processor based supercomputer which was found useful not only for solving problems on which BARC scientists are working, but also by the aeronautical engineers for simulating aerodynamics of certain new designs of aeroplanes.

Nondestructive testing and quality assurance in manufacturing processes are important activities in a nuclear power programme. Expertise developed by some of the BARC engineers and scientists in those areas are of such a high standard that other industries also call upon BARC for getting specialised service in these areas.

The motto of BARC is to support good work which is relevant to the nuclear programme, which includes power generation, utilisation of research reactors and use of radio isotopes; and to encourage research of high quality even if it is not directly related to the nuclear programme. Relevance and excellence are the key words in performance evaluation.

M R Balakrishnan, Head, Library & Information Division
Bhabha Atomic Research Centre, Mumbai (Bombay) 400085 India.
Email:mrbala@magnum.barc.ernet.in

⇦ Bhabha Discussing the Design of the Apsara Reactor Control

R D Choksi, J R D Tata, Homi Bhabha and Pandit Nehru ⇨

Acknowledgements

Resonance gratefully acknowledges the help received from Jayant Rao.

"I had the idea that after the war I would accept a job in a good university in Europe or America, because universities like Cambridge or Princeton provide an atmosphere which no place in India provides at the moment. But in the last two years, I have come more and more to the view that provided proper appreciation and financial support are forthcoming, it is one's duty to stay in one's own country and build up schools comparable with those that other countries are fortunate in possessing."

Homi J Bhabha in a letter to Sir Sorab Saklatvala dated 12 March 1944.

"For every thousand scientists who can do reasonably good work in a good scientific atmosphere, there is only one who can create the atmosphere for himself in a place where it does not exist, and this alone is a test of the outstanding scientist"

Homi J Bhabha in a letter to B M Udgaonkar dated 8 July 1963.

Homi J Bhabha (1909–1966), sculptor B Vithal, courtesy Tata Institute of Fundamental Research, Mumbai. *Also see Article-in-a-box on pp.3–5.*

KRN-A 326 *Resonance* – July 1998 ISSN 0971-8044

Homi Jehangir Bhabha

(1909 – 1966)

Registered with Registrar of Newspapers in India vide Regn. No. 66273/96.

Tholasi Prints India